Summer 8-15-2012

A Ranksum Statistics Based Framework to Decipher Transcription Regulation

Iuan-bor Chen
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences

Computational Biology

Dissertation Examination Committee
Patrick Jay, chair
Jason Mills
Rakesh Nagarajan
Jeanne Nerbonne
Gary Stormo
David Wilson

A Ranksum Statistics Based Framework to Decipher Transcription Regulation

by
Iuan-bor Chen

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2012
Saint Louis, Missouri

Abstract

The unbiased generation of specific and meaningful hypotheses from the deluge of data generated by modern genomic methods remains a challenge. These datasets require increasing level of expertise to analyze fully, and are often underutilized even in the originating lab. It would be desirable to have a computational strategy that is easy to implement, robust against outliers and missing data, and broadly applicable to diverse experimental designs. In this dissertation, I present a set of ranksum statistics-based analytical methods as a framework to extract testable hypotheses from large and complex datasets. To illustrate its utility, this framework was applied to two clinically important biological questions. In both instances, my method yielded novel molecular mechanisms that were subsequently validated with both *in vitro* and *in vivo* experiments. In the first study, gene expression profiles from multiple mouse models of cardiac hypertrophy were analyzed to reveal a novel interaction between transcription factors Nkx2-5 and Egr1, providing mechanistic insight into how *Nkx2-5* haploinsufficiency leads to exacerbated cardiac hypertrophy and poor survival in these mice. In the second study, thousands of microarray samples acquired from public data repositories were analyzed to quantitatively define tissue-specific expression pattern for every gene represented on a microarray platform. The tissue-specific expression data was then used to identify novel transcriptional regulators of brown fat gene expression program in adipocytes. The successful application of the analytical framework in these examples, regardless of their differing experimental design, highlights its adaptability in facilitating discoveries in a wide array of biological problems.

Acknowledgments

The thesis work presented here would not be possible without the support of my supervisor and good friend Patrick Jay. Pat is the very model of a modern physician-scientist. He is tireless and fearless in pursuit of ever-grander scientific challenges. I can't thank Pat enough for being the perfect guide and saving me from many pitfalls through the course of my thesis work. On a personal note, I wish to thank Pat and his wife Kathleen for treating me like family. They have shown me warmth and generosity that will not be soon forgotten. The great team Pat has assembled in the lab over the years is another reason I feel very fortunate to call the Jay lab my scientific home. Suk Regmi, Vinay Rathi, Diana DeAndrade, Yali Lu, and Min Li have chipped in countless hours of experiments and discussion that propelled me towards the finish line. They are my friends and my teachers, and they will always have my gratitude. I also wish to thank my current and past thesis committee members Drs. Gary Stormo, Jean Nerbonne, Rakesh Nagarajan, Jason Mills, David Wilson, Daniel Kelly, and Weixiong Zhang, for invaluable guidance and infinite patience they have provided me with.

Washington University has been an exceptionally generous institution. I have received financial support from the Department of Computational Biology, the Department of Pediatrics, and the Lucile P. Markey Pathway, which was led by Drs. Daniel Link and Alan Schwartz. Labs of Drs. Daniel Kelly, Anthony Muslim, Jeff Milbrandt, David Wilson, Jean Nerbonne, Paul Hruz, Jason Mills, Joshua Rubin, Jennifer Duncan, David Rudnick, and Andreas Burkhalter have each shared their equipment, reagents and expertise with me in my thesis research. I also owe my gratitude to the administrative staff at the Division of Biology & Biomedical Sciences, the Office for International Students and Scholars, and the Department of Pediatrics,

especially Melanie Puhar and Barbara Kloeckener, who magically made all the paperwork invisible.

Graduate study would have been a very lonely road without the camaraderie of fellow DBBS graduate students. Drs. and future Drs. Christian Perez, Christina Chen, Thomas Kraft, Edward Esparza, Ashwin Desikan, Enrique Ramos, Liwei Cheng, Yat Tang, Yue Yun, KC Yang, Bin Zhang, Aneeza Salim, JW Feng, Claire Schulkey and Julia Honold have my eternal gratitude for our collaborations and friendship. I want to also thank many good friends St. Louis has blessed me with. I thank Anatoly, Raj, Bryan, Antti, Goska, Ivan, Daniel, Riddhi, Mayur, Matt, Nailun, Tricia, Youngin, Jennifer, Kristine, German, Katia, Christine, Carmen, Simone, Anni, Sara, Lena, Conny, Barb, Romy, Lina, Lisa, Chuck, Anja, AK, Jessi, Verena, and John for sticking with me through good and not-so-good times alike.

Through my thesis work, my parents in Taiwan and my family in Canada have been unbelievably encouraging. I feel incredibly lucky and grateful to have them. I am also fortunate to have my brother Andy as a colleague who always shows up in my lab at the right time with exactly the right advice. Lastly, I thank my wife Maria; she has always believed in me and made everything seem possible.

Table of Content

List of Tables

List of Figures

Abbreviations

36B4	ribosomal protein, large, P0 (Rplp0)
4-OHT	4-hydroxytamoxifen
ACE	angiotensin I converting enzyme 1
Acta1	actin, alpha 1, skeletal muscle
Actc1	actin, alpha, cardiac muscle 1
Ascl1	achaete-scute complex homolog 1
ATP	adenosine triphosphate
αMHC	myosin, heavy chain 6, cardiac muscle, alpha (Myh6)
ASD	atrial septal defect
BAT	brown adipose tissue
BMI	body mass index
CCL2	chemokine (C-C motif) ligand 2
Cebpa	CCAAT/enhancer binding protein (C/EBP), alpha
CHD	congenital heart disease
ChIP	chromatin immunoprecipitation
Cm	whole-cell membrane capacitance
CnA	protein phosphatase 3, catalytic subunit/calcineurin A
DMSO	dimethyl sulfoxide
Egf	epidermal growth factor
Egr1	early growth response 1
Elovl5	ELOVL fatty acid elongase 5
Esrrg	estrogen-related receptor gamma
Gata4	GATA binding protein 4
GEO	Gene Expression Omnibus
GFP	green fluorescent protein
GO	the Gene Ontology
GWAS	genome-wide association study
HDAC	histone deacetylase
IGF1	insulin-like growth factor 1
IPTG	isopropyl β-D-1-thiogalactopyranoside
MBP	maltose binding protein
Mef2	myocyte enhancer factor 2
Myh6	myosin, heavy chain 6, cardiac muscle, alpha
Myh7	myosin, heavy chain 7, cardiac muscle, beta
Myt1l	myelin transcription factor 1-like
Nab1	NGFI-A binding protein 2

Nab2	NGFI-A binding protein 2
NFAT	nuclear factor of activated T-cells
Ntg	non-transgenic
Nkx2-5	NK2 transcription factor related, locus 5 (Drosophila)
PET	positron emission tomography
Pou3f2	POU class 3 homeobox 2
Pparg	peroxisome proliferator-activated receptor gamma
Prdm16	PR domain containing 16
qPCR	quantitative real time polymerase chain reaction
Rplp0	ribosomal protein, large, P0
RT	reverse transcription
shRNA	small hairpin RNA
Srebf1	sterol regulatory element binding transcription factor 1
TAC	transverse aortic constriction
TAG	triacylglycerol
TFBS	transcription factor binding sequence/site
Tg	transgenic
Tpm1	tropomyosin 1 (alpha)
Ucp1	uncoupling protein 1
VSD	ventricular septal defect
WAT	white adipose tissue

Chapter 1. Introduction

Obesity

Obesity is an epidemic

Obesity is a condition where excessive body fat accumulation negatively impacts a person's health. The most commonly used measure of obesity is based on the body mass index (BMI), which is a ratio between a subject's weight in kilograms and the square of the subject's height in meters, although an excess of body mass may not be attributed to an excess of body fat in every case. The center for Disease Control and Prevention defines a BMI between 25 and 30 as overweight, and a BMI greater than 30 as obese. The 2007-2008 National Health and Nutrition Examination Survey, which sampled from the entire U.S. population [1], placed the prevalence of obesity at 33.8% and the combined prevalence of overweight and obesity at 68%. Since the inception of the survey in the 1960's, the prevalence of obesity has been on a steady linear incline, and has since then doubled. In the same period, the prevalence of obesity in children and adolescents has tripled to 16.9% in 2008. This is an alarming trend since obese children tend to become obese adults. Obesity is a risk factor in a plethora of diseases including coronary heart disease, hypertension, stroke, and type 2 diabetes, and is a major contributor to mortality. In the overweight and the obese, it is estimated that each 5kg/m2 increase BMI translates to a 30% increase in all-cause mortality [2]. On average, a BMI increase from 24 to 32 leads to a loss of three years in life expectancy. The health care cost attributed to obesity is estimated to be $147 billion per year in 2008 [3].

Causes and Treatments of Obesity

In animals, unspent energy intake is stored as triacylglycerols (TAGs) in the adipose tissue as an evolutionary solution to the scarcity of food in the environment. Obesity, therefore, is a consequence of both the unbalanced caloric homeostasis, and the individual's response to the unbalance as dictated by genetic and environmental variables. Rare monogenic causes of obesity exist, such as defects in genes involved in the *leptin* signaling pathway which regulates appetite. In the majority of cases, obesity is a complex condition associated with multiple contributing factors. The prevalence of obesity differs in subpopulations defined by ethnicity, gender, age, and socio-economic status [1,4]. Studies following twins reared apart reported only slightly lowered correlation in BMI between the siblings compared to twins reared together, suggesting strong genetic influences [5,6]. A contrasting study, however, showed that obesity can spread among non-related members in social networks, potentially through shared environment and behavior [7]. Recent papers showed the composition of gut flora is altered in obese mice compared to the lean mice [8,9]. While gut flora composition and the BMI of the host have an interdependent and complex relationship, these observations nevertheless open the possibility that obesity may be transmitted between individuals in this fashion. Currently, options to treat obesity are limited. Drugs which suppress appetite or fat absorption have questionable efficacy. Surgical interventions are dangerous and available only to the most severe cases. Lifestyle modification, which is the most effective and safe measure for the prevention and treatment of obesity, is difficult to implement. Novel therapeutic strategies that overcome the drawbacks in existing therapies are therefore desired.

White adipose tissue is a regulatory organ of metabolism

In obese individuals, the increase in adiposity occurs mainly in white adipose

tissues (WAT). In the body, WAT is distributed in discrete depots that are different from each other in composition and properties [10,11]. WAT consists mainly of white adipocytes and little connective tissues. White adipocytes each feature a single large lipid droplet where excess calories are stored as TAGs. In addition to being the energy reservoir in the body, WAT is increasingly recognized as an endocrine organ. The adipocytes produce a number of cytokines and chemokines, as well as other secreted factors known as adipokines [12]. Adipokines such as leptin and adiponectin are regulators of energy intake and metabolism. Leptin, encoded by the *ob* gene, relays satiety signal from WAT to the brain to suppress appetite [13]. Adipokines have also been implicated in disease processes. Adiponectin, for example, is protective against type 2 diabetes [14]. In obesity, adiposity increases to cause hypertrophy in the adipocytes which precedes eventual adipocyte hyperplasia. The pathology of obesity, however, is not merely the quantitative increase in WAT fat content. Another hallmark of obesity is the dysregulation of secretory factors. The serum level of adiponectin, for example, is inversely proportional to body fat percentage [14]. Obese WAT additionally up-regulates chemokines such CCL2 [15]. Increased secretion of CCL2 leads to macrophage infiltration in WAT, resulting in the inflammation process which is another qualitative changes occurring in obese WAT.

Brown fat as treatment for obesity

Brown adipose tissue (BAT) shares the ability to uptake lipids with WAT but is structurally and functionally distinct. BAT acquires its appearance from the abundance of mitochondria in brown adipocytes. Uncoupling protein 1 (Ucp1), a BAT-specific protein, is uniquely expressed in these mitochondria to uncouple the electron transport from ATP production to generate heat. In this regard, brown adipocytes resemble cardiomyocytes in that both are metabolically active organ with

high mitochondrial content. The former converts chemical energy from fat into thermo energy while the latter into mechanical work. Small mammals rely on BAT to maintain thermo homeostasis as they are incapable of shivering. Human infants also possess appreciable amount of BAT estimated at 5% of body weight, most of which is lost by adulthood. Several recent discoveries have casted the BAT as a potential therapy for obesity. First, PET scan on workers frequently exposed cold environment showed increased BAT mass; suggesting the endogenous BAT may be activated in adults. Second, dispersed brown fat clusters are found in the WAT depots in mice subjected to cold-exposure or β-stimulation. Whether these 'brite' adipocytes originate from of white adipocytes or circulating stem cells is undetermined, they are distinct from brown adipocytes found in BAT depots. Third, recent works have delineated much of the gene regulatory networks required to specify brown adipocyte cell fate from the common progenitor cells shared with skeletal myocytes. In particular, the transcription regulator Prdm16 is sufficient to drive the brown fat gene program in a variety of cells. It is estimated that activation of half of the endogenous BAT complement in a typical adult can expend energy equivalent to 4 kilogram of body fat per year. Chemical or genetic interventions designed to activate endogenous BAT or BAT-like clusters in WAT would thus be a powerful tool to combat the obesity epidemic.

Heart Failure

Heart failure and intervention

Heart failure is a serious condition where the heart is unable to pump sufficient amount of blood through the systemic circulation to meet the needs of the body. Heart failure represents a major public health challenge both domestically and abroad. Longitudinal data collected since the 1960's have shown a steady increase in the

prevalence of heart failure, which is currently estimated to affect 5.7 million in the U.S [16] and 22 million worldwide, despite continuing improvement in heart failure therapeutics. Ironically, the increases in the prevalence of heart failure have been attributed to advances in treatments for other cardiovascular diseases since these patients can now survive with their conditions long enough to develop heart failure. Current standard of care for treating heart failure patients includes use of ACE inhibitor and β-blockers for those in the early stages of the disease. In the most severe cases, surgical interventions such as implantation of left ventricular assist devices or cardiac transplantations are needed. Despite their remarkable efficacy in prolonging patients' lives, modern heart failure treatments merely slow the progression of the disease but cannot halt or reverse the pathogenic processes. As a result, prognosis for heart failure patients remains poor. In the four decades following the first availability of the aforementioned treatments in the 1960's, the survival rate at five years after the initial diagnosis only improved from 70% to 59% [17], which is still worse than most forms of cancer [18]. Although a large body of knowledge describing the molecular basis of heart failure has accumulated, it is likely that other important pathogenic pathways remain untargeted by the existing treatments. Research effort to identify previously unknown pathogenic pathways is therefore critical in the development of novel therapeutic strategies that would overcome limitations of current treatments for heart failure.

Cardiac hypertrophy is an integral part in the pathogenesis of heart failure

Heart failure is a complex condition with multiple etiologies. Major risk factors found in large epidemiological studies include male sex, age, diabetes, and obesity. Heart failure is often the terminal progression of many cardiovascular conditions. Hypertension, ischemic heart disease, cardiomyopathy, and valvular disease can all

lead to heart failure. Cardiac hypertrophy, or the enlargement of the heart, is a condition frequently associated with these diseases, and is often an integral part of the pathogenesis in the development to heart failure. The heart is a muscular organ with 90% of its mass made up by cardiomyoctyes. As in all muscle, the heart undergoes hypertrophic response in reaction to increased load. Cardiac hypertrophy can be divided into two categories depending on the underlying cause. Physiologic hypertrophy follows stimuli such as exercise training and pregnancy, is reversible and does not progress to heart failure. In contrast, pathologic hypertrophy results from cardiac insults such myocardial infarction, cardiomyopathy, and hypertension. In these cases the contractile property of the myocardium is impaired by the underlying conditions, and the heart must compensate to maintain cardiac output via hypertrophic response. Cardiac hypertrophy, in turn, leads to further diastolic and systolic abnormalities that degrade cardiac output. As a result, pathologic hypertrophy does not usually undergo remission and progresses into heart failure if left unmanaged.

Molecular basis of cardiac hypertrophy

Cardiac hypertrophy in response to pathological stimuli involves a complex network of molecular sensors, signal transducers, transcription factors, and a host of downstream target genes (reviewed in [19]). Hypertrophic signals such as paracrine growth factors and physical stretching of the myocytes are sensed by cell surface receptors. These include integrin, a transducer of mechanical stress, endothelin-1 receptor, adrenergic receptor, EGF receptor, IGF1 receptor, and thyroid hormone receptor. These receptors are targets of some of the most successful drugs to treat heart failure such as ACE inhibitors and β-blockers. Nuclear receptors, such as the thyroid hormone receptor, can translocate into the nucleus upon activation. Signals from other receptors rely on signal transduction pathways to reach the transcription

factors in the nucleus. These pathways often begin with the disassociation of activated g-proteins from the receptors. The g-proteins, in turn, pass the signal through a network of kinases to the transcriptional and translational machineries to produce the hypertrophic response. Intracellular calcium is another critical component of the signal transduction leading to cardiac hypertrophy. Calcineurin, for example, is a phosphatase that is activated by sustained increase in the intracellular calcium level [20,21]. When activated, calcineurin dephosporylates a number of proteins, including transcription factor NFAT. In the nucleus, activated transcription regulators such as NFAT, Mef2, and HDACs are recruited to modulate the expression of target genes that induces various aspects of the hypertrophic response. These may include switching from fatty acids to glucose as fuel [22], reactivation of fetal genes [23], increased protein synthesis [24], increased apoptosis [25,26], and cell growth.

Relevant mouse models of cardiac hypertrophy

Transverse aortic constriction model (TAC) Since its invention in 1991[27], transverse aortic constriction in rodents has been a widely adapted model of pressure overload induced cardiac hypertrophy. In this model, a suture band is placed around the transverse aorta using microsurgical techniques. The degree of aortic banding is carefully controlled to produce a stenosis that results in a pressure gradient averaging around 40mmHg, doubling the normal range, in the aorta proximal to the suture band [27]. This increases the afterload that the heart must pump against. The heart mitigates the increased afterload either by increasing the ventricular wall thickness or by increasing the ventricular diameter in accordance with the Law of LaPlace [28]. TAC is highly effective in inducing cardiac hypertrophy. Typically, differential gene expressions in the TAC mice compared to that in the sham operated littermates can be measured mere hours following the operation while increased cardiac mass can be

observed less than a week post operation. TAC, however, is not without drawbacks. An experienced animal surgeon is required to maintain high consistency and low mortality in TAC operations.

Calcineurin overexpression model *Calcineurin* transgenic (*CnA*-Tg) mice[21] express a constitutively active form of calcineurin driven by a αMHC promoter, thereby producing a high level of activated CnA protein in the heart. Calcineurin, as described earlier, catalyzes the dephosphorylation of cytoplasmic NFAT in response to sustained increases in intracellular calcium level. Dephosphorylated NFAT translocates to the nucleus and activates a transcription program that is sufficient to cause cardiac hypertrophy. *CnA*-Tg mice thus develop extensive cardiac hypertrophy as early as five weeks of age. Similarly to the TAC model, the *CnA* overexpression model is widely used in studies of cardiac hypertrophy. The *CnA* overexpression model offers several advantages over the alternative. Since it is a genetic model, there is a high degree of consistency in the cardiac hypertrophy developed in the *CnA*-Tg mice. The extent of cardiac hypertrophy in these mice is often greater than what can be achieved in surgical models. Lastly, there are no specialized staff and instrumentation requirements to utilize the *CnA*-Tg mice in a study.

Role of *Nkx2-5* in cardiac hypertrophy and heart failure

Nkx2-5 is a cardiac-specific transcription factor required for normal heart development. Mutations of human *NKX2-5* are associated with inherited congenital heart defects (CHDs) [29,30]. In mice, homozygous deletion of *Nkx2-5* halts heart development and is lethal to the embryo at E9-10. Heterozygous deletion of *Nkx2-5* (*Nkx2-5*$^{+/-}$) leads to a less severe phenotype. *Nkx2-5*$^{+/-}$ mice in C57 background survive birth but carry a risk for conduction defects and CHDs [31]. In the embryo, Nkx2-5 expression first appears in the progenitor cells (in the heart field) specifying

cardiac cell fate. Nkx2-5 expression persists through heart development and into the postnatal adult heart [32]. There is little literature describing the role Nkx2-5 plays in the adult heart. Nkx2-5 is up-regulated in animal models of cardiac hypertrophy [33] and, presumably, is involved in the development of the hypertrophic response. Experimental evidence, however, does not converge on a unified paradigm of how Nkx2-5 modifies the hypertrophic process. Transgenesis of dominant negative Nkx2-5 in the adult mice leads to myocyte degeneration and increases the severity of the detrimental effect of injected toxins on the heart [34,35]. Similarly, conditional deletion of Nkx2-5 in post-mitotic mouse hearts results in conduction and contraction defects in the heart [36], although the severity of the phenotype decreases as the age of the mice increases when the deletion is triggered. In contrast, overexpression of wild-type Nkx2-5 using a αMHC promoter leads to the development of a conduction defect and heart failure [37,38]. These conflicting results suggest that the role for Nkx2-5 in the adult heart may be highly context-dependent. Further experiments where Nkx2-5 level is minimally perturbed may be necessary to disambiguate the function of Nkx2-5 in the postnatal heart.

Transcriptional regulation and expression profiling

Transcriptional is a fundamental mechanism underlying the causality of an organism's inherent genotype and its observed phenotype. At the simplest level, transcription factors occupy transcription factor binding sequences (TFBSs) upstream of a given gene to promote mRNA synthesis by basal transcriptional machinery. Several mechanisms to modulate gene expression arise even within such a simple model of transcription. Protein-DNA interactions between transcription factors and TFBSs are governed by thermodynamics. As a result, variations in TFBS sequence identity directly affect the dynamics of transcription factor-DNA complex formation.

Other properties of TFBSs, such as distance, orientation, and phase on the double helix relative to a transcription start site can each influence interactions between transcription factors and transcriptional machinery. In addition, transcription factors may exert activating or suppressing effect. In mammalian cells, transcriptional regulation is far more complex than the above model. A typical mammalian gene is served by many transcriptional factors whose TFBSs are found anywhere from the introns of the gene to hundreds of kilobases away from the transcriptional start. Promoter/enhancer sequences can be covalently modified to vary their availability to transcription factors. Transcription factors are also subject to positive or negative regulation from upstream regulators, co-factors, or ligands, either by physical interaction or chemical modification. Furthermore, nascent mRNAs can be stabilized or degraded via covalent modification or microRNA regulation. Together these mechanisms form a regulator network to ensure robust temporally- and spatially-specific expression of genes at appropriate levels.

In normal development, as well as in pathogenesis of diseases, genes are differentially regulated through modifications in transcriptional regulatory networks. High-throughput experimental methods such as oligonucleotide microarray and, more recently, parallel sequencing of cDNA (RNA-Seq), have been developed in the past 15 years to simultaneously quantify large number of mRNA transcripts. The discipline of bioinformatics was created to manage datasets from these experiments where the raw output has grown from mere megabytes, in the early generations of microarrays, to hundreds of gigabytes in the latest RNA-Seq platform. As these methods became accessible and increasingly ubiquitous in non-specialized laboratories, a gap between the ability to generate data and the expertise to analyze

data has widened. Massive amount of data are deposited in open-access public data repositories such as Gene Expression Omnibus but publications resulting from meta-analyses of public datasets have been few in comparison. A computational framework that is easy to implement, robust against outliers, insensitive to variations in data processing practices by originating labs, and flexible to accommodate addition of orthogonal data would promote utilization of high-throughput datasets and maximize value of research investment.

Chapter 2: Gene discovery by non-parametric analysis of massively pooled, public expression microarray data

Iuan-bor D. Chen[1], Vinay K. Rathi[1], Diana S. DeAndrade[1], Maria Efimova[1], Patrick Y. Jay[1,2]

Departments of [1]Pediatrics and [2]Genetics
Washington University School of Medicine
Box 8208
660 South Euclid Avenue
St. Louis MO 63110

Correspondence to PYJ.
Phone: 314-362-2174
Fax: 314-286-2892
E-mail: jay_p@kids.wustl.edu

ABSTRACT

The vast trove of expression profiles in the Gene Expression Omnibus (GEO) contains biological gems, but their mining poses formidable challenges not least because data normalization is a complex problem. We present a simple, non-parametric solution. Datasets based on the same microarray platform are classified into two or more experimental groups. No pre-processsing of expression data is necessary except for ranking probesets within each microarray in order of signal intensity. The ranks of a probeset are then ranked across the microarrays that comprise the experimental groups. Rank sum tests determine relative differences in gene expression. Applied to the thousands of profiles of mouse tissues in GEO, the method could define tissue-specific gene expression patterns across a wide dynamic range. This permitted the accurate prediction of genes specifically relevant to the function of an organ as well as the physiologic functions of an organ. Reasoning that genes that perform a function are more specifically expressed in tissues that perform

the function than other tissues, we designed bioinformatic screens for transcription factors that regulate brown adipocyte gene expression. Comparing the heart and brown adipose tissue (BAT), the two most mitochondrial-rich tissues in the body, to white adipose tissue (WAT), we predicted and validated a novel function for the transcription factor Esrrg in the induction of BAT genes in white fat. Because the heart and other tissues that express Esrrg do not express BAT genes, we hypothesized that an adipocyte cofactor is required with Esrrg for BAT gene expression. By comparing WAT and BAT to the heart, brain, kidney and skeletal muscle, we discovered that an isoform of the transcription factor Srebf1 could induce BAT genes in a muscle cell culture model. The results illustrate how our computational strategy can make genomic data mining fruitful and more accessible to a larger community.

AUTHOR SUMMARY

Microarrays are a popular means to quantify the expression of thousands of genes in a sample simultaneously. An experiment generates more data than an investigator could ever interpret. The Gene Expression Omnibus (GEO) was hence established so that others could glean additional insights from functional genomic data. Data mining, however, typically requires a computational infrastructure that few laboratories have because the preparation of heterogeneous datasets for conventional, parametric statistical tests is complex. Given the huge number of GEO profiles for virtually every cell and organ in the body, we show that a non-parametric statistical approach is simpler and yet sufficiently powered to discover genes and biological patterns. In proof-of-concept analyses we identified many genes known to be specifically expressed in various organs and relevant to their function. Furthermore, we could accurately predict the physiologic functions of an organ based on its list of genes ranked by their specificity of expression. To prove the utility of our method

prospectively, we designed bioinformatic screens and discovered two transcription factors that regulate gene expression in brown fat. The more accessible computational strategy can thus streamline the notoriously labor-intensive process of gene discovery by harnessing the collective effort of the biomedical research community.

INTRODUCTION

Transcriptional profiling with microarrays is a popular means to explore gene expression patterns. Consequently, a massive amount of data exists for virtually all the tissues in the body. Submission of the data to public databases like the Gene Expression Omnibus (GEO) is promoted so that secondary analyses can increase the scientific return on a collectively huge investment.[39] Compelling discoveries support this intent,[40-42] but typically only the contributors ever analyze the data. Non-standardized data filtering, scaling and normalization can hinder the replication of the original analytical results. Differences in sample preparation, hybridization, scanning, and image processing can confound the reproducibility of an experiment, which complicates the pooling of data across studies.[43,44]

All these issues are problematic because conventional statistical analyses of microarray data are parametric. Data normalization, which can be vexing, is necessary to assure a uniform probability distribution. We reasoned that a non-parametric approach based on rank sum statistics could facilitate secondary analyses of aggregated microarray datasets. Several properties of rank sum statistics make it ideal to circumvent difficult problems with microarray data. First, the test is insensitive to the underlying distribution of the observations, obviating the need for normalization. Second, the results are robust against outliers and missing data because the test omits some of the variance information contained in the datasets. The resulting loss of statistical power can be offset by large sample sizes. Finally, rank sum statistics

requires relatively little computational power and is amenable to parallelization.

To demonstrate the utility of the approach, we defined the tissue-specificity of gene expression patterns across the genome for multiple organs by a meta-analysis of GEO datasets based upon the Affymetrix Mouse Genome 430A 2.0 (MG430Av2) microarray. In the past, genome-wide, tissue-specific gene expression patterns have been determined by counting ESTs, which biases against rare transcripts.[45,46] More recently, tissue-specific gene expression patterns have been surveyed using a few microarrays per organ, which limits the power to detect differences.[47-51] In contrast, the results of the rank sum method suggest that the approach is robust and sensitive across a wide dynamic range of expression because the analysis draws from up to hundreds of samples per organ and thousands for comparison between tissues.

A useful product of the method is the ranking of genes from most specifically to most counter-specifically expressed in an organ. The term "counter-specific" is used instead of "repressed" or "suppressed" so as not to imply an active regulatory mechanism. Organs that perform a common function likely employ the same genes. The genes are probably more specific to organs that perform the function than elsewhere. This premise logically suggests bioinformatic screens for genes based on threshold criteria for tissue specificity in various organs. The threshold criteria are determined by set logic, prior knowledge, or both.

We thus developed bioinformatic screens for genes that regulate brown adipocyte gene expression. Under normal conditions, uncoupling protein 1 (*Ucp1*) is expressed predominantly in brown adipose tissue (BAT) but can be induced in the white adipose tissue (WAT) of animals exposed to the cold.[52-54] Ucp1 uncouples mitochondrial oxidative phosphorylation to generate heat instead of ATP. BAT shares the abilities with WAT for fat uptake and with cardiac and skeletal muscle for energy

production via abundant mitochondria. The transcriptional coactivator, *Prdm16*, is essential for the specification of BAT from a pre-myogenic precursor in the embryo and the induction of BAT in adult WAT.[55-59] Other genes are known to regulate BAT gene expression.[60] Additional genes might be found by comparing tissue-specific gene expression patterns in WAT, BAT and mitochondrial-rich organs like the heart. Through a series of bioinformatic screens and biologic experiments we identified estrogen-related receptor gamma (Esrrg) and sterol regulatory element binding transcription factor 1 (Srebf1) as transcriptional regulators of *Ucp1* and other BAT markers.

RESULTS

A non-parametric algorithm ranks genes by their tissue specificity of expression

Although expression microarrays are perceived to be insensitive to differences among genes expressed at low levels, the non-parametric method could rank the tissue-specificity of gene expression across a wide dynamic range. First, the *fluorescence rank* of a gene, as represented by a probeset, was determined in order of decreasing brightness for each microarray in the pooled dataset. Second, the *expression rank* of a gene was determined by ranking the fluorescence ranks across all the microarrays, which in turn were classified as being from the tissue of interest or not. Third, we performed rank sum tests of the expression ranks for every gene in the two microarray groups. The analysis generates a Z-score that lends itself to genome-wide ranking of genes (Supplementary Table 2.1). Negative or positive Z-scores correspond to genes specifically or counter-specifically expressed in the tissue compared to all others. The *tissue-specificity rank* of each gene derives from ranking the Z-scores. The analyses were performed on datasets from a single microarray platform so that cross-platform differences in probeset performance would

not confound the fluorescence rank relationships between genes (Fig. 2.1). Examples of genes that are specific, counter-specific or neither in the heart follow.

The transcription factor *Nkx2-5* and cardiac actin isoform *Actc1* are expressed almost exclusively in the heart.[32,61] Consistent with their respectively low and high levels of expression, their median fluorescence ranks are 3542 and 15 among the 45101 probesets in 175 profiles of postnatal murine hearts and 25449 and 25485 in 2270 non-heart arrays ($P < 1 \times 10^{-101}$ for both genes; Fig. 2.1b). From these data were calculated the specificity ranks. *Nkx2-5* and *Actc1* are ranked 89 and 11 in the heart but much lower in non-heart tissues (Fig. 2.1c).

At the opposite extreme, the fatty acid elongation protein *Elovl5* is expressed at a much lower level in the heart than elsewhere.[62] The median fluorescence ranks of the gene are 4485 and 1404 in the heart and non-heart arrays ($P < 1 \times 10^{-101}$; Fig. 2.1b). The tissue-specificity rank of *Elovl5* in the heart is 44970. The Z-score is large and positive, consistent with counter-specific expression of the gene in the heart (Fig. 2.1c).

The ribosomal subunit *Rplp0*, also known as Arbp or 36B4, is expressed at comparable levels in diverse tissues.[63] The median fluorescence ranks of *Rplp0* were 179 and 221 in the heart and non-heart arrays ($P = 1$; Fig. 2.1b). The gene has a tissue-specificity Z-score near zero in most organs, reflecting its ubiquitous and consistent level of its expression (Fig. 2.1c).

Rarely is a gene expressed exclusively in a single organ. A limited pattern of expression can still be relatively tissue-specific. The algorithm provides a means to judge relative specificity, as illustrated by the comparison of actin isoforms expressed in skeletal muscle. Skeletal alpha-actin, Acta1, comprises ≥95% of the sarcomeric actin in skeletal muscle, whereas Actc1 is a minor fraction.[61] *Acta1* and *Actc1* have

specificity ranks of 68 and 1031 in skeletal muscle, but *Actc1* is still more specifically expressed in muscle than other tissues except the heart ($P < 1 \times 10^{-34}$; Fig. 2.1c).

Genome-wide rankings of tissue-specific expression offer insight into gene and organ function

The expression pattern of a gene can offer insight into its function. Several observations suggest that our algorithm could offer similar insights. First, the tissue-specificity rank of a gene in each organ appears to correlate well with a function specific to the organ, as illustrated by the 10 transcription factors deemed most specific in 15 tissues (Fig. 2.2). On average, mutation or knockout of ~6 of the 10 transcription factors is associated with a human disease or mouse phenotype related to the function of the gene in the tissue. Of course, tissue-specific genes that do not have a known mutant phenotype could still have a function. For example, myelin transcription factor 1-like (*Myt1l*) is the third most brain-specific transcription factor and with two other transcription factors can convert fibroblasts directly into neurons ($P = 4 \times 10^{-274}$; Fig. 2.2, Supplementary Table 2.1). [64] The two, *Pou3f2* and *Ascl1*, are also highly brain-specific ($P < 1 \times 10^{-107}$ and $< 1 \times 10^{-17}$. The least significant statistic of the two probesets that represent each gene is presented.)

The repertoire of genes expressed by an organ determines its functions, which suggests that the tissue-specificity ranks of genes could be adapted to predict organ functions. Based on the Gene Ontology (GO) annotations for each gene, we performed a rank sum analysis of biological processes according to their distribution on the tissue-specificity rank list of an organ. Perusal of the analyses shows that the specialized functions of an organ can be predicted from pooled microarray data (Fig. 2.3, Supplementary Table 2.2). Closer inspection suggests that additional insights may be gleaned. For example, the testis and ovary produce gametes and sex hormones.

Consistent with their shared functions, there is substantial overlap of the statistically significant genes and biologic processes in the two gonadal organs (Fig. 2.2, 2.3). The analyses can even suggest less obviously shared genes or functions. Genes and GO terms related to immunity are expected for the spleen and thymus, but they are also enriched in the intestine, lung and skin, thus highlighting their common function as a barrier against pathogens (Fig. 2.3, Supplementary Tables 2.1 and 2.2).

Esrrg regulates brown adipocyte gene expression in white fat

The prior analyses of pooled microarray data revealed known tissue-specific genes and biologic processes. We next tested predictive utility based upon the premise that tissues that perform a common function utilize the same genes that are more specifically expressed in said tissues. We developed a bioinformatic screen for genes that could induce brown adipocyte gene expression in white adipose tissue while considering two key facts. First, *Prdm16* is a critical determinant of the brown fat phenotype.[59] Second, the heart and BAT are the most mitochondrial-rich tissues in the body, whereas WAT is mitochondrial poor. We defined genes whose expression is more or as specific to BAT and heart and more or as counter-specific to WAT than *Prdm16* in each tissue. The screen yielded 104 probesets, representing 97 unique genes (Fig. 2.4a, Supplementary Table 2.3). The set is enriched for genes whose products localize to the mitochondria (51% versus 9.3% of all genes on the microarray, $P = 3.07 \times 10^{-16}$; Fig. 2.4b). In addition, 11% of the genes are regulated directly or indirectly by the transcription factor Esrrg.[65] which is also in the set, whereas <1% of all represented genes are ($P = 6.38 \times 10^{-7}$). The increased mRNA expression of *Esrrg* in BAT relative to WAT was verified by quantitative RT-PCR (Fig. 2.4c).

These observations led us to examine the ability of Esrrg to induce mRNA

markers of brown fat in OP9 cells, a white adipocyte cell culture model.[66] Transiently transfected Esrrg could up-regulate the expression of the BAT markers *Ucp1*, *Cidea*, and Pcg-1α in differentiated OP9 adipocytes, while down-regulating the WAT marker resistin (Fig. 2.4d, 2.4e). Esrrg had no effect in undifferentiated OP9 pre-adipocytes, suggesting the requirement for an adipocyte-specific co-factor (Fig. 2.4d).

BAT gene expression can be induced in white adipocytes by the Pparg agonist rosiglitazone.[34-36] Rosiglitazone similarly induced BAT markers as well as *Esrrg* in differentiated OP9 adipocytes (Fig. 2.4f). To determine whether induction depends upon *Esrrg*, we applied 4-hydroxytamoxifen (4-OHT), an Esrrg antagonist.[67,68] The drug suppressed the rosiglitazone-mediated induction of *Ucp1* and *Cidea* (Fig. 2.4f). *Esrrg* was likewise suppressed, which suggests either that Esrrg regulates its own expression in a positive feedback loop or that 4-OHT acts on another target that regulates *Esrrg* expression. Unfortunately, efforts to transfect OP9 cells with shRNA constructs, as an alternative means to knock down Esrrg activity, were unsuccessful.

Mice subjected to the cold induce BAT genes in WAT. After housing for three days at 4^{o}C, *Ucp1* and other BAT markers were up-regulated up to ten-fold in the WAT. The up-regulation was abrogated by 4-OHT, which is consistent with a role for *Esrrg* in the induction of BAT in WAT (Fig. 2.5a). Visual inspection during gross dissection suggested that there was more intrascapular BAT after cold exposure, but the expression of *Ucp1* and other markers was only slightly upregulated or unchanged (Fig. 2.5b). 4-OHT may not inhibit *Ucp1* expression in BAT because *Esrrg* is expressed an order of magnitude higher than in WAT (Fig. 2.4c). Treated mice maintained normal body temperature while housed in the cold, consistent with the maintenance of Ucp1 expression in BAT.

Srebf1 induces brown adipocyte gene expression in cultured myoblasts

Esrrg induces BAT gene expression in adipocytes, but *Esrrg* is also highly expressed in mitochondrial-rich tissues that have no Ucp1 (Fig. 2.6a.). White adipocytes may possess a cofactor that renders them competent to express BAT markers in the presence of Esrrg. A hypothetical cofactor should be expressed more specifically in white and brown adipose tissue than mitochondrial-rich organs. *Prdm16* is probably not the cofactor, as the brain, heart and kidney express the gene, albeit at lower levels than in BAT (Fig. 2.6a). A bioinformatic screen was designed for genes that have negative tissue-specific Z-scores in WAT and BAT and positive Z-scores in brain, heart, muscle and kidney. The screen yielded 614 annotated probesets representing 564 unique genes, including 53 transcriptional regulators (Supplementary Table 2.4). *Srebf1* was selected for further analysis in a muscle cell culture model because its expression pattern closely resembled the criteria for the hypothesized Esrrg co-factor (Figure 2.5b). Of two alternatively spliced forms, Srebp-1a and -1c, Srebp-1c was chosen for transfection because it is the predominant isoform in white and brown adipocytes.[69]

Transfection of C2C12 myoblasts with Srebp-1c significantly induced the expression of *Ucp1*, other BAT markers and *Esrrg* (Fig. 2.6c). If Srebp-1c and Esrrg act together to induce brown adipocyte gene expression, then inhibition of Esrrg should repress the induction in C2C12 cells by Srebp-1c. Consistent with this hypothesis, 4-OHT abrogated the up-regulation of BAT markers by Srebp-1c, just as it did in OP9 cells and in the WAT of cold-exposed mice. RNAi knockdown of Esrrg likewise repressed the induction of BAT markers by Srebp-1c, whereas a control shRNA had no effect. Srebp-1c had no effect on the expression of the myocyte-specific marker tropomyosin 1 alpha (*Tpm1*) (Fig. 2.6c).

DISCUSSION

Nature does not readily reveal the inner workings of life. Biologists expend great effort on screens to discover genes and experiments to elaborate gene function. We suggest that in some cases the total effort might be distributed across the larger community. Expression microarrays have become a commonplace tool, whereas statistical analyses seem increasingly complex. This asymmetry fosters the perception that data mining is accessible only to laboratories that have substantial bioinformatics support. Some complex problems do require sophisticated solutions, but a simple, non-parametric statistical approach to massively pooled, heterogeneous datasets uncovers a fertile middle ground. We apply the approach to provide a database that ranks the tissue-specificity of gene expression in multiple organs and evidence for the function of two transcription factors in BAT gene expression.

Microarray data are typically normalized to remove variation that is not of biological interest. Normalization remains a challenging problem even when an experiment is well-defined.[41,70] Normalization of heterogeneous, pooled datasets is immensely harder. Non-parametric statistics is suited for cases where the underlying distribution of data is uncertain and when large sample sizes are available. Despite the greater statistical power of parametric tests, rank sum tests applied to aggregated GEO datasets could clearly define tissue-specific gene expression patterns in a variety of organs. The large sample sizes afforded by GEO ensure the robustness of analyses against random and technical variation. Genes are detected across a wide dynamic range of expression, as shown by the identification of low-expressed transcription factors with organ-specific functions. Higher order biological patterns hidden in GEO are also detectable, as shown by the prediction of the specialized functions of organs.

In addition to demonstrating the utility of the strategy, the tissue-specificity rank

lists for multiple organs may be useful in their own regard. For example, given a chromosomal interval determined by genetic linkage or association, one may ask whether any gene within the locus is expressed in an organ relevant to the phenotype. More broadly, the statistical strategy should be amenable to any biological question that can be addressed by comparison of microarray samples classified into two or more groups. The resulting list of ranked genes can inform experimental priorities, as shown by implication of *Esrrg* and *Srebf1* in BAT gene expression.

A good screen should be simple and fruitful. Mechanistic questions remain regarding the role of *Esrrg* and *Srebf1* in the determination of the BAT phenotype, but the experimental results prospectively demonstrate the yield of a bioinformatic strategy based upon simple statistical analyses of public microarray data. Other genes that were identified in the same screens as *Esrrg* and *Srebf1* likely have relevant functions. For example, Pgc-1α was identified in the screen for genes specific to the heart and BAT but not WAT. Originally cloned from a brown fat cDNA library, Pgc-1α is essential to mitochondrial biogenesis and BAT thermogenesis.[71,72] In the screen for genes specific to BAT and WAT but not other mitochondrial-rich tissues, *Tle3* was represented by two probesets along with *Srebf1* and 51 other transcriptional regulators. *Tle3* was recently discovered in a screen of 18,292 cDNA vectors to co-regulate adipogenesis with *Pparg*. Interestingly, *Cebpa*, a positive control in the cDNA screen, was likewise identified in the bioinformatic screen.[73]

Obviously, no screen is perfect or all-encompassing. Sensitivity is limited by the design of the screen. For example, the present analyses would not identify ubiquitously expressed genes that regulate BAT gene expression. Specificity is a concern too, but seemingly false positive results could be genes that have functions implicitly included in a bioinformatic screen. Hence, the screen that yielded Esrrg

captured genes that perform mitochondrial functions, which is consistent with the high abundance of mitochondria in the heart and BAT. Despite the limitations, we suggest that they are outweighed by the relative simplicity of the non-parametric statistical approach and the consequent ease of extracting biological insights from GEO. Tapping into the collective labor that populated the public database, a research team could spread the herculean effort otherwise expended by them to elucidate a gene or biological function.

METHODS

Data acquisition and processing

Via the GEO interface (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL1261), we downloaded 4983 Affymetrix Mouse Genome 430A 2.0 microarray datasets (MG430Av2; Affymetrix, Inc., Santa Clara, CA; GEO platform accession GPL1261) in SOFT format. A custom Perl script parsed the SOFT file and assigned a source organ to each sample by matching a set of keywords for each organ to the sample source and the submitter's description fields. Computer-generated organ designations were manually confirmed by IDC and PYJ. An overview of the microarray collection is shown (Fig. 2.1a). Of the 4983 microarrays in the collection, 2538 were excluded because the biological samples were of unknown origin, fetal in origin, cultured cell lines, or derivatized, e.g., a subpopulation of cells purified from an organ, resulting in a set of 2445 microarray datasets used in the tissue-specific gene expression analysis. Probeset annotations (release 30) in .csv format were downloaded from the Affymetrix NetAffx portal (www.affymetrix.com).

Statistical algorithm

To detect differential gene expression between two groups consisting of a total of *m* samples run on the same microarray platform, the meta-analysis algorithm first converted absolute fluorescence values (i.e., VALUE column from the SOFT formatted file) within each microarray sample into *fluorescence ranks* from 1 to *n* in order of decreasing brightness, where *n* is the number of probesets represented on the microarray (Fig. 2.1b). For samples with fewer than *n* fluorescence values reported, the fluorescence ranks were scaled to range from 1 to *n*. For each of the *n* probesets, the algorithm ranked the fluorescence rank values in descending order across the *m* samples to determine *expression ranks* from 1 to *m*. A Wilcoxon rank sum test was then applied to the expression rank values for the two groups being compared to generate a Z-score, where a greater absolute Z-score represents more significant differential expression between the two groups being compared. The sign of the Z-score indicates which of the two groups has higher relative expression of the gene as measured by the probeset. To assess tissue-specific expression, the samples from the tissue of interest are compared to all others that are not. The *tissue-specificity rank* of a gene was determined by ranking Z-scores generated by the rank sum tests from the most negative to the most positive; a lower Z-score reflects greater tissue specificity.

Gene Ontology analysis

Gene Ontology (GO) biological process terms associated with the probesets on the Affymetrix MG430Av2 microarray were extracted from the probeset annotation file. The tissue-specificity of a given GO term was determined by a rank sum test. Probesets from a tissue-specificity rank list for a particular organ were divided into two groups that were either annotated with the given GO term or not. The enrichment

of the annotated group toward the top or the bottom of the ranklist was described by the rank sum test Z-score and p-value. The enrichment of GO terms in a set of probesets identified in the bioinformatic screen for potential brown adipose tissue regulators was evaluated with GeneMerge.[74] Significance thresholds were set at a false discovery rate of 5%.

Cell culture

The OP9 pre-adipocyte cell line was a generous gift from Dr. Paul Hruz (Washington University School of Medicine, St. Louis, MO). The OP9 culture was maintained and differentiated into adipocyte-like cells as previously described.[66] The OP9 pre-adipocytes were transfected using Optifect (Invitrogen, Carlsbad, CA) as per the manufacturer's instructions. Transfected OP9 adipocytes were cultured for an additional three days after transfection and differentiation before harvesting. In pharmacologic experiments, differentiated OP9 cells were incubated with the Pparg agonist rosiglitazone (10 μM in DMSO, Sigma Aldrich, St. Louis, MO) and/or Esrrg antagonist 4-hydroxytamoxifen (4-OHT; 10 μM in DMSO, Sigma Aldrich) for 48 hours before harvesting.

The C2C12 skeletal myoblast cell line was purchased from and maintained according to the protocol supplied by ATCC (Manassas, VA). C2C12 cells were transfected using Lipofectamine 2000 (Invitrogen) per the manufacturer's instructions. The cells were maintained for an additional 36 hours after transient transfection before harvesting. In pharmacologic experiments, C2C12 cells were incubated with 4-OHT at 10 μM for 48 hours before harvesting.

Expression Plasmids

A pcDNA3.1(-)-Met-Esrrg expression construct was a gift from Daniel Kelly (Burnham Institute for Medical Research, Lake Nona, FL). A pSV SPORT Srebp-1c

expression construct was purchased from Addgene (Cambridge, MA). Esrrg targeting (Target gene 26381) and control shRNA constructs from the RNAi Consortium library (Broad Institute, Cambridge, MA) were obtained via The Genome Institute at Washington University and the Children's Discovery Institute (St. Louis, MO).

Animal Studies

12-week-old female FVB/N mice were treated with intraperitoneal injection of 4-OHT (25 mg/mL in sunflower seed oil and 10% DMSO) or vehicle once daily at 0.1 mg/g/day. The mice were placed in a 4°C room six hours after the first injection. They received food and water ad libitum. The core temperature of each mouse was measured twice daily with a K-type thermocouple probe digital thermometer. Abdominal white adipose tissue and intrascapular brown adipose tissue were harvested after 72 hours of cold exposure. The experiments were approved by the animal studies committee at Washington University School of Medicine.

RNA Isolation and Quantitative PCR

RNA was isolated from the tissue samples using Trizol (Invitrogen), according to the manufacturer's protocol. The precipitated RNA was resuspended in 175 µL of RNA lysis buffer from the SV Total RNA Isolation Kit (Promega, Madison, WI) and purified according to the manufacturer's protocol. Total RNA from cultured cells were isolated with SV Total RNA Isolation Kit only. cDNA was synthesized from 2.5 µg of total RNA using SuperScript VILO cDNA Synthesis Kit (Invitrogen) according to manufacturer's protocol.

Quantitative PCR of cDNA was performed in Cycler iQ system (Bio-Rad, Hercules, CA) or Mx3005P QPCR System (Stratagene, Santa Clara, CA) in a 40 µL reaction containing iQ SYBR Green Supermix (Bio-Rad) and PCR primers at a final concentration of 200 nM. Primer sequences were designed with the PerlPrimer or

obtained from PrimerBank.[75,76] Primers were purchased from Sigma Aldrich or Integrated DNA Technologies (Coralville, IA). The primer sequences are available in the supplemental methods. Relative mRNA expression was calculated using the ΔΔ-Ct method and normalized to Hprt1.

Student's t-tests were used to compare groups in the qPCR studies. The data are reported as mean ± s.e.m.

ACKNOWLEDGEMENTS

We thank Anatoly Tzekov for technical advice on OP9 cell culture. We thank Drs. Alan Schwartz and David Wilson for helpful comments.

FIGURES

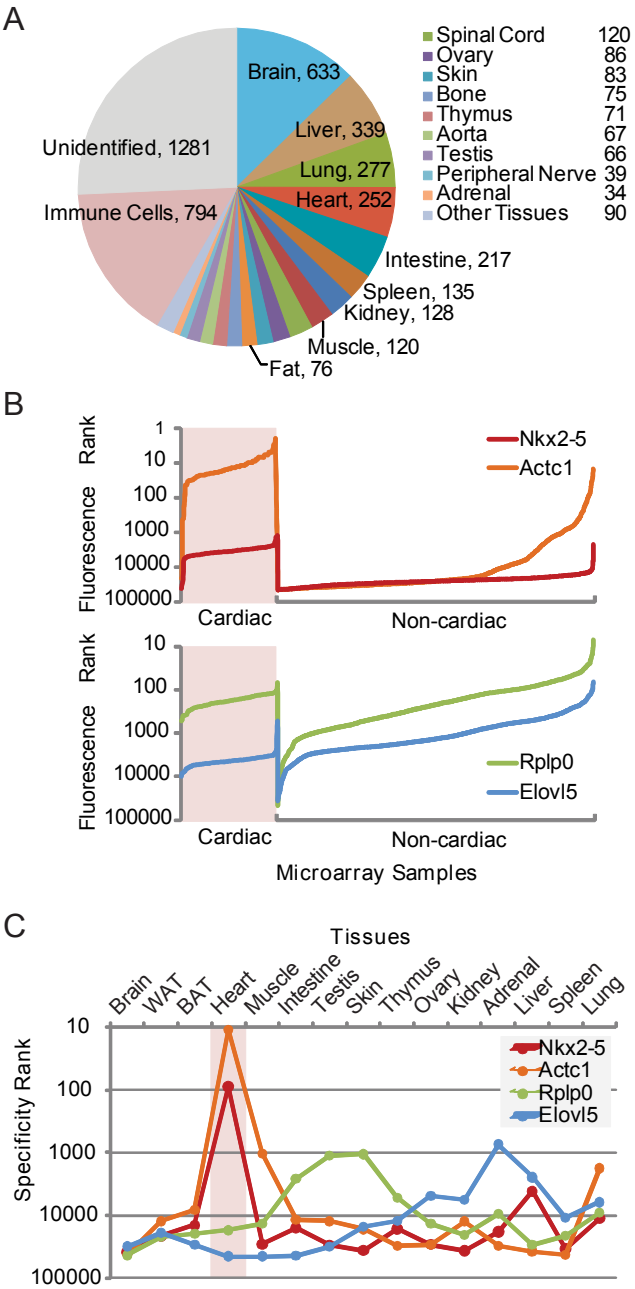


Figure 2.1: A non-parametric statistical approach defines tissue-specific gene expression patterns from massively pooled microarray datasets. (a) The tissue of origin for the 4983 Affymetrix MG430Av2 microarray samples downloaded from GEO and the number of samples for each tissue used in subsequent analyses are shown. (b) A gene, represented by a probeset, is ranked by its fluorescence intensity on a microarray for every sample in the collection. Top, fluorescence ranks for the

cardiac-specific transcription factor *Nkx2-5* (red tracing) and cardiac actin isoform *Actc1* (orange) are shown. Their fluorescence ranks are generally higher in cardiac microarrays, consistent with their cardiac specificity. Bottom, fluorescence ranks for the ribosomal subunit *Rplp0* (green) and fatty acid elongation protein *Elovl5* (blue) are shown. The fluorescence ranks of *Rplp0* are equivalent in cardiac and non-cardiac samples, which is consistent with its known uniform expression in diverse tissues and its use as a reference or housekeeping gene for mRNA quantification. *Elovl5* is expressed at low levels in the heart and likewise has generally lower fluorescence ranks in cardiac samples. The fluorescence ranks are plotted in reverse logarithmic scale. The non-cardiac portion of the tracings is compressed horizontally to facilitate the interpretation of the plots. (c) The tissue-specificity ranks of *Nkx2-5*, *Actc1*, *Rplp0*, and *Elovl5* are shown for various organs in reverse logarithmic scale; higher points denote greater specificity.

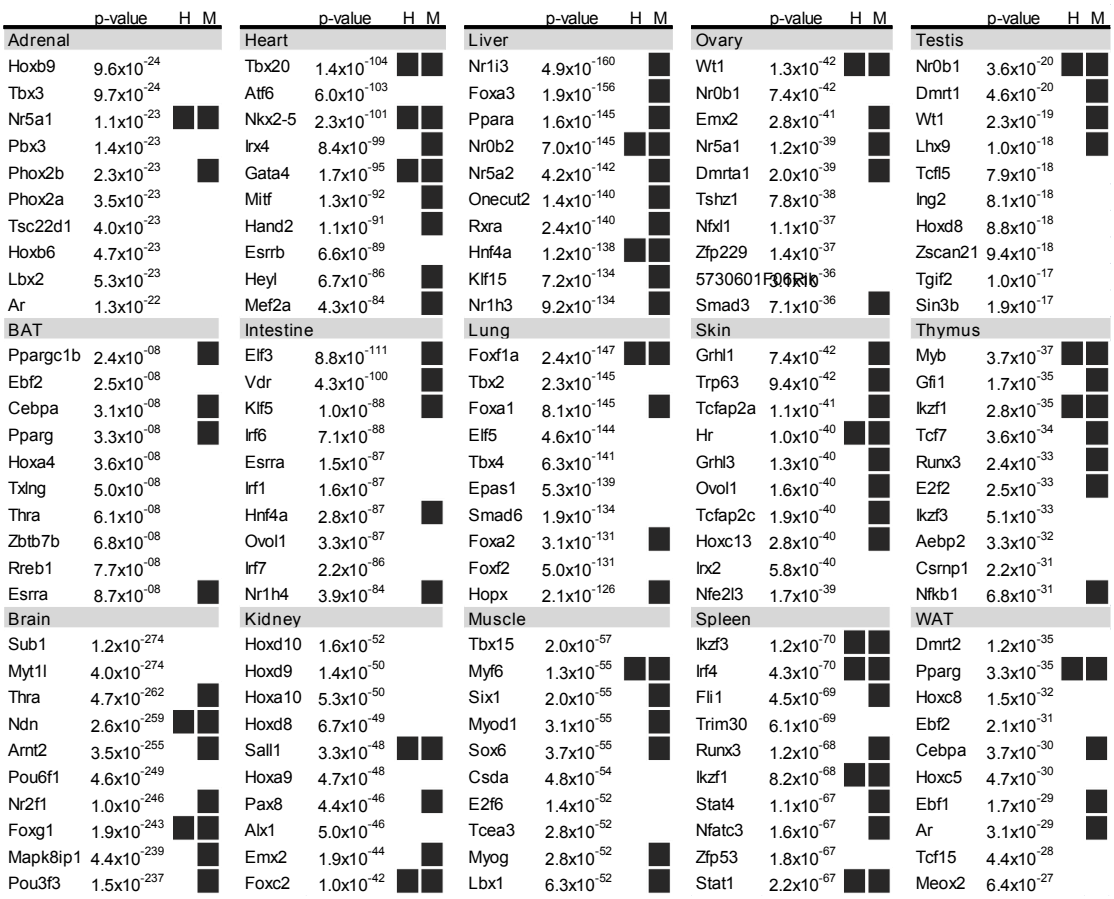


Figure 2.2: The ten most specifically expressed transcription factors in 15 different tissues are listed. If several probesets represent transcription factor, the most significant probeset is shown. Genes were identified as transcription factors if their associated Gene Ontology terms included "0006355 regulation of transcription, DNA-dependent" or "0006350 transcription" and "0003677 DNA binding". The official gene symbols and p-values derived from our tissue specificity rank sum analysis are shown. Filled boxes indicate the existence of human (H) or mouse (M) mutant phenotypes related to the function of the transcription factor in the organ.

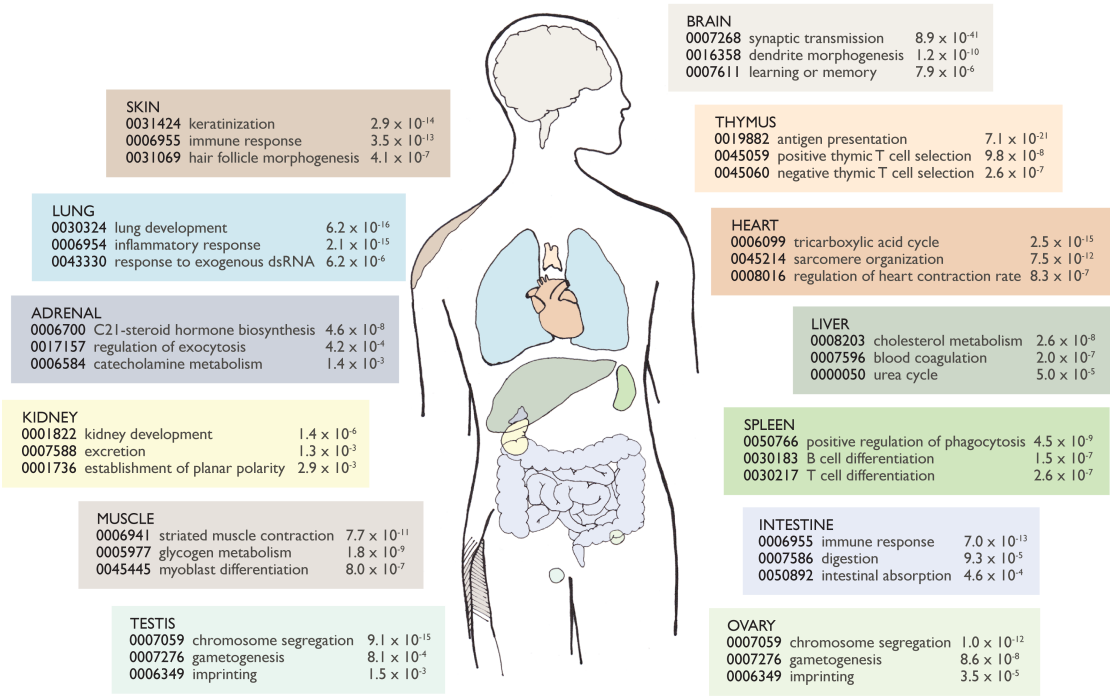
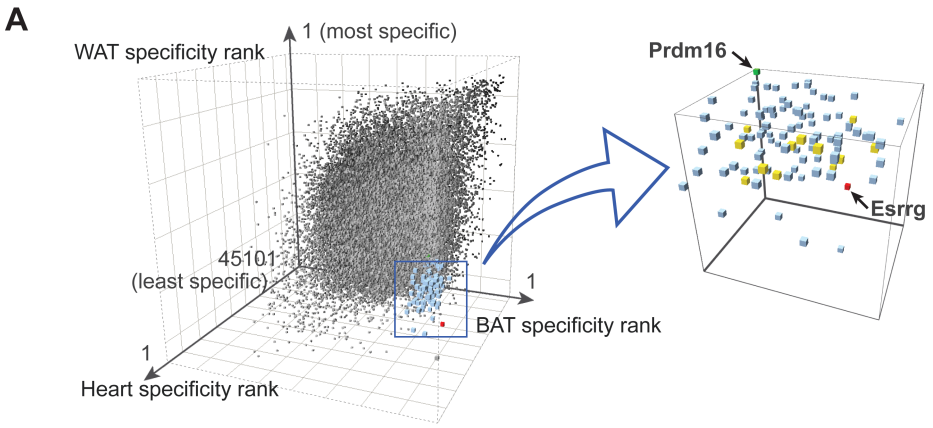


Figure 2.3: The function of an organ can be predicted from the GO biological process terms assigned to genes ranked by their specificity of expression in the organ. Representative, significant functions for a number of organs are shown here with their GO accession number and P-value. Significance thresholds were defined by setting a false discovery rate to 5%. The full lists of GO terms are provided in the Supplementary Table 2.2.



B

Accession	Description	p-value	Background fraction	Study fraction	Enrichment
MF:0004129	cytochrome-c oxidase activity	4.77×10^{-03}	22/14184	3/56	35X
BP:0009060	aerobic respiration	5.62×10^{-03}	4/12812	2/54	119X
BP:0007005	mitochondrion organization	8.69×10^{-03}	32/12812	3/54	22X
CC:0005739	mitochondrion	3.07×10^{-16}	1355/14495	34/67	5X
CC:0005743	mitochondrial inner membrane	1.02×10^{-07}	268/14495	12/67	10X
CC:0042645	mitochondrial nucleoid	3.49×10^{-03}	23/14495	3/67	28X
Esrrg response genes (Alaynick et al., 2007)		6.38×10^{-07}	164/21815	9/79	15X

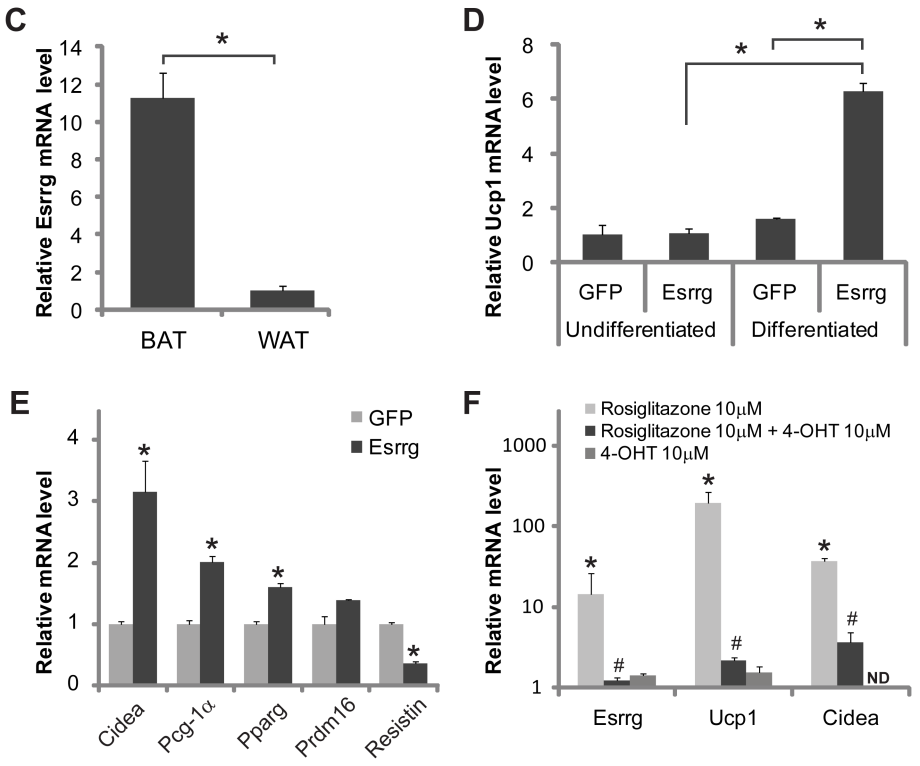


Figure 2.4: Esrrg induces BAT gene expression program in white adipocytes. (a) Schematic representation of the bioinformatic screen. Left, the tissue specificity scores in BAT, WAT, and the heart are plotted for all 45101 probesets (gray) represented on the MG430Av2 microarray. The probesets (blue) that are more or as

specific to BAT and the heart and more or as counter-specific to WAT than *Prdm16* were identified. Right, the probesets are shown with *Prdm16*, *Esrrg*, and *Esrrg*-regulated genes in green, red, and yellow, respectively. (b) The subset is significantly enriched for mitochondrial genes and genes regulated by *Esrrg*. Significance thresholds were set by a false discovery rate of 5%. (c) Quantitative RT-PCR analysis of *Esrrg* mRNA in murine BAT and WAT (*, $P = 0.023$; $n = 4$). (d) Esrrg induces *Ucp1* mRNA expression in differentiated but not undifferentiated OP9 cells; GFP control transfections are shown for comparison (*, $P < 0.05$; $n = 3$). (e) Essrg induces additional BAT markers, *Cidea*, Pgc-1α, *Pparg*, and *Prdm16*, and suppresses the WAT marker Resistin in differentiated OP9 cells. (*, $P < 0.05$; $n = 3$). (f) The Pparg agonist rosiglitazone induces BAT markers in differentiated OP9 cells, an effect that is suppressed by the Esrrg antagonist 4-hydroxytamoxifen (4-OHT). Data are reported relative to vehicle-treated cells. (*, $P < 0.05$ versus vehicle-treated cells; #, $P < 0.05$ versus rosiglitazone-treated cells; $n = 4$). Error bars indicate s.e.m. BP, biological process. CC, cellular component. MF, molecular function. ND, not detected.

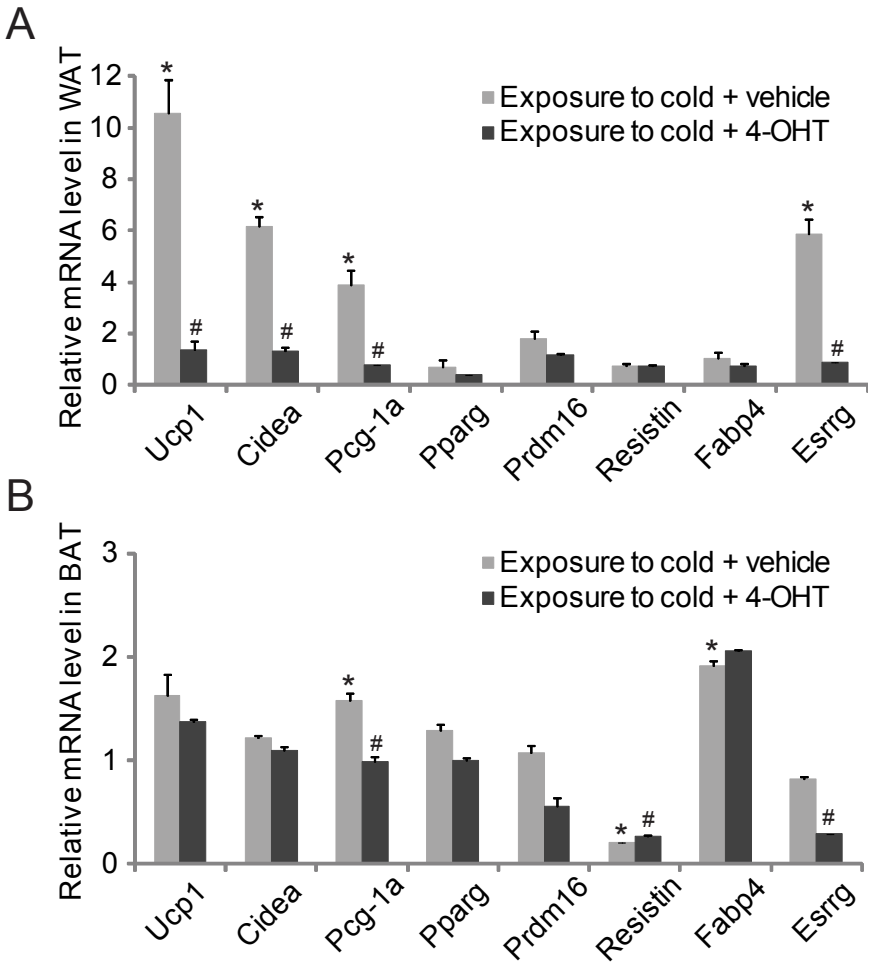


Figure 2.5: (a) Exposure of mice to the cold induces brown adipocyte gene expression in WAT. The Esrrg antagonist 4-OHT suppressed the effect of cold exposure. (b) 4-OHT did not suppress the cold-induced induction of Ucp1 in BAT possibly because Esrrg is expressed at higher levels in BAT than WAT. Fold changes are stated relative to the WAT or BAT of vehicle-treated mice housed at room temperature. (*, $P < 0.05$ with respect to vehicle treated mice at room temperature; #, $P < 0.05$ with respect to vehicle-treated mice exposed to the cold.) $n = 6$ animals per group. Error bars indicate s.e.m.

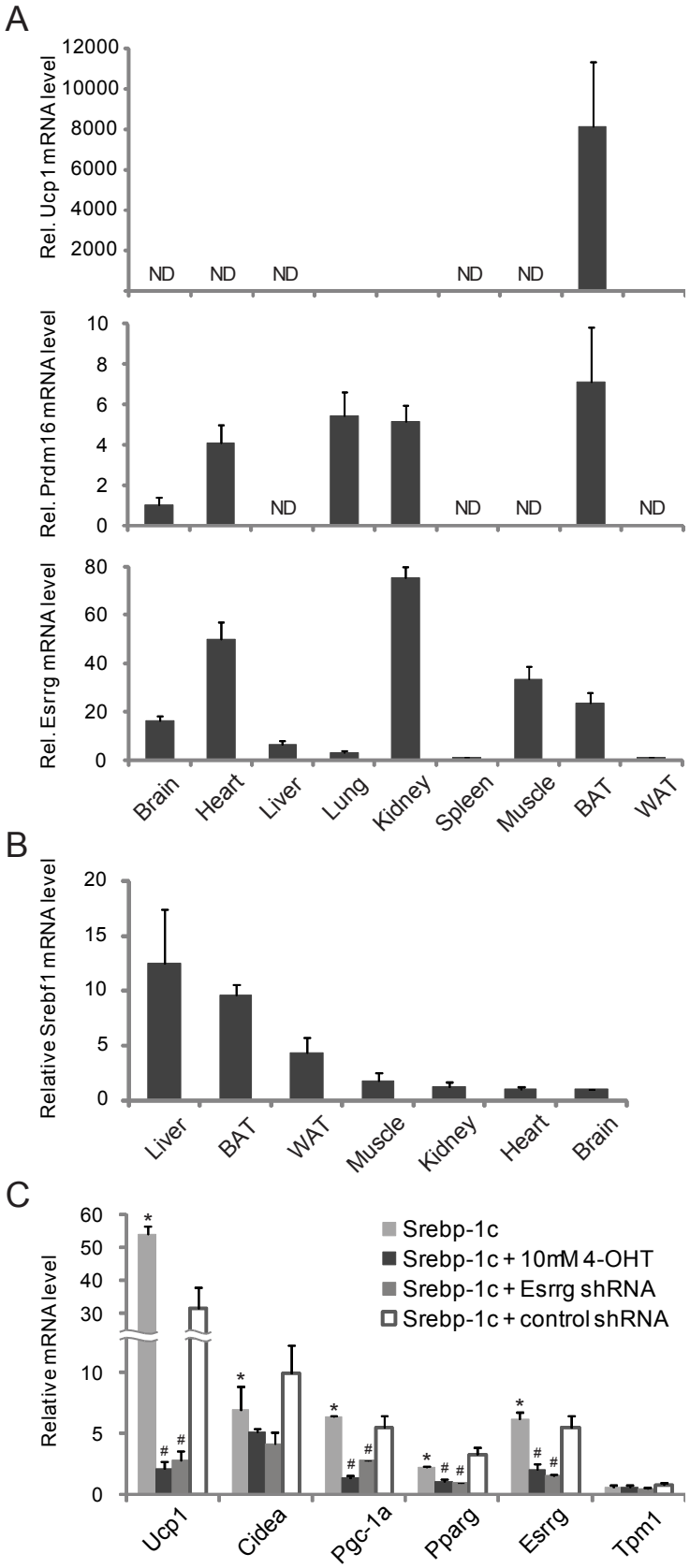


Figure 2.6: (a) *Ucp1* is specifically expressed in BAT, whereas *Prdm16* and *Esrrg* are expressed in various other tissues, suggesting that an adipocyte cofactor is necessary

for *Esrrg* to induce the brown adipocyte gene expression program ($n = 4$; ND, not detected). (b) *Srebf1* is more highly expressed in the liver, WAT and BAT than other mitochondrial rich tissues ($n = 4$). (c) Srebp-1c induces the brown adipocyte gene expression program in C2C12 myoblasts. Induction depends upon endogenous Esrrg, as shown by inhibition or knockdown of Esrrg by 4-OHT or shRNA, respectively. A lacZ control shRNA construct had no effect on Srebp-1c induction of brown fat markers. The measurements are normalized to GFP-transfected cells. Transfections were performed in triplicate. Error bars indicate s.e.m. (*, $P < 0.05$ compared to GFP-transfected controls; #, $P < 0.05$ compared to Srebp-1c-transfected cells; $n = 3$).

Chapter 3: Nkx2-5 physically interacts with Egr1 to modulate pathological cardiac hypertrophy

Iuan-bor D. Chen[1], Kai-Chien Yang[2], Suk D. Regmi[1], Vinay K. Rathi[1], Jeanne M. Nerbonne[2], Patrick Y. Jay[1,3]

The Departments of [1]Pediatrics, [2]Developmental Biology, and [3]Genetics

Washington University School of Medicine

Box 8208

660 South Euclid Avenue

St. Louis, MO 63110

Please address correspondence to: Patrick Y. Jay, Phone: (314) 362-2174; Fax: (314) 286-2892; E-mail: jay_p@kids.wustl.edu

Running title: Nkx2-5-Egr1 interaction modulates pathologic cardiac hypertrophy

Key Words:

Nkx2-5, Egr1, Nab1, Nab2, cardiac hypertrophy, computational biology, TRANSFAC, Calcineurin, transverse aortic constriction, haploinsufficiency

Background: *Nkx2-5* is critical in fetal heart development but the postnatal role for *Nkx2-5* is less well known.

Results: *Nkx2-5*$^{+/-}$ mice exhibit aggravated pathology and poor survival in cardiac hypertrophy.

Conclusion: Nkx2-5 interacts with hypertrophy mediator Egr1 and suppresses the activation of Egr1 gene targets in cardiac hypertrophy.

Significance: A novel pathogenic mechanism in cardiac hypertrophy is identified, demonstrating the utility of the bioinformatics presented.

SUMMARY

The cardiac-specific transcription factor *Nkx2-5* is critical in embryonic heart development. In mice, *Nkx2-5* null mutation is embryonically lethal. *Nkx2-5* heterozygous (*Nkx2-5*$^{+/-}$) animals survive birth but have high rate of congenital heart defects. The postnatal role for *Nkx2-5* is less well understood. The present study aims to delineate how *Nkx2-5* functions in the adult heart by characterizing *Nkx2-5*$^{+/-}$ mice in the context of pathological cardiac hypertrophy. *Nkx2-5*$^{+/-}$ pups born without heart malformations were otherwise healthy and indistinguishable from wild-type littermates. In these mice, hypertrophic stimuli such as *calcineurin* overexpression (*CnA*-Tg) and transverse aortic constriction led to exacerbated hypertrophic responses and poor survival compared with wild-type littermates. Computational analyses in the proximal promoters of genes differentially regulated both in cardiac hypertrophy and *Nkx2-5* haploinsufficiency under hypertrophic conditions revealed an enrichment of transcription factor binding sites for hypertrophy mediator early growth response 1 (Egr1). *In vitro*, Nkx2-5 physically interacted with Egr1 and Egr1-specific suppressor

NGFI-A binding protein 1 (Nab1) to down-regulate Egr1-dependent transcriptional activation. *NGFI-A binding protein 2* (*Nab2*) is a *Nab1* paralog and a target of this regulation. An increase in Nab2 level reduced Nkx2-5- and Nab1-dependent suppression of Egr1 activity. *In vivo*, decreased *Nab2* expression in *Nab2*$^{+/-}$/*CnA*-Tg mice attenuated the hypertrophic response and improved survival in comparison to *Nab2* WT/*CnA*-Tg littermates. Thus, *Nkx2-5* can modulate cardiac hypertrophy through suppressing Egr1-dependent transcriptional activation in target genes such as *Nab2*. The computational method presented here is useful for discovering transcriptional regulatory pathways in physiological or pathological processes involving differential gene regulation.

INTRODUCTION

Pathological cardiac hypertrophy is an important sequela in heart conditions of diverse etiologies. In response to detrimental stimuli such as pressure overload, gene programs are activated in the heart to up-regulate protein synthesis, leading to increases in cardiomyocyte volume and cardiac mass. Cardiac hypertrophy is initially a compensatory mechanism but can progress to heart failure over time if left unmitigated [77]. Transcriptional regulation is a critical mechanism in the pathogenesis of cardiac hypertrophy. The re-activation of genes normally restricted to the developing fetal heart, such as *atrial natriuretic factor (ANF)* and *β-myosin heavy chain (Myh7)*, is a universal hallmark in pathological hypertrophic process [78]. Nkx2-5 is a cardiac-specific transcription factor with essential functions in cardiac development. Nkx2-5 has been shown to drive the expression of *ANF*, *Myh7*, and other cardiac hypertrophy marker genes [79], and is itself up-regulated in adrenergic-induced cardiac hypertrophy [33]. In a study of heterozygous human *NKX2-5* mutation, several individuals haploinsufficient for *Nkx2-5* presented with idiopathic

left ventricular hypertrophy [30]. This led us to wonder whether the reduced *NKX2-5* level in affected individuals might lead to hypersensitivity to biomechanical stimuli that induce cardiac hypertrophy. We tested this hypothesis by characterizing the cardiac hypertrophic response in *Nkx2-5*$^{+/-}$ mice challenged with pathological surgical and genetic stimuli.

Nkx2-5 exhibits a gene-dosage dependent developmental phenotype in the embryonic heart. In mice, a germline, homozygous knockout mutation in *Nkx2-5* leads to the arrest of cardiac development at the looped heart tube stage [79,80] and is embryonically lethal. Mice with a heterozygous deletion of *Nkx2-5* (*Nkx2-5*$^{+/-}$) survive birth, but the mutation is associated with an increased incidence of congenital heart defects (CHDs) such as atrial and ventricular septal defects (ASDs and VSDs) [31]. Similarly, heterozygous mutations of *NKX2-5* are associated with anatomic and conduction defects in human patients [30,81]. *Nkx2-5* expression persists into adulthood [32], but the postnatal role of *Nkx2-5* is less well defined. Because the genetic deletion of *Nkx2-5* is embryonic lethal, alternative experimental strategies have been employed to elucidate the postnatal functions of *Nkx2-5* [34,82]. Conditional deletion of *Nkx2-5* leads to degeneration in the conduction and contractile functions in the perinatal mouse model [36,82], suggesting that *Nkx2-5* may be involved in the maintenance of cardiac function in healthy cardiomyocytes. The functional role of *Nkx2-5* in pathological conditions is equally ambiguous. Overexpression studies of *Nkx2-5* in mice have yielded mixed results, ranging from a hypertrophy-free phenotype to heart failure as early as 4 months of age [37,38]. Transgenic mice expressing a dominant negative form of Nkx2-5, however, demonstrate both cardiac conduction defects and increased cardiac apoptosis when challenged with doxorubicin [34,35], suggesting that *Nkx2-5* may be protective in the adult heart. The available experimental data thus

point to the involvement of *Nkx2-5* in the pathogenesis of cardiac hypertrophy, although the precise role of *Nkx2-5* in the development of cardiac hypertrophy was not examined directly.

The *Nkx2-5*$^{+/-}$ mice expresses half the normal levels of *Nkx2-5* mRNA and protein [81], which is a minimal perturbance in comparison to the alteration of mRNA levels seen in traditional overexpression and knock-out models. The characterization of the *Nkx2-5* haploinsufficiency model should therefore lead to a more physiologically relevant model of how *Nkx2-5* functions in the postnatal heart and how it affects the pathogenesis of cardiac hypertrophy. The present study identifies a putative mechanism underlying the *Nkx2-5* dosage-dependent modulation of cardiac hypertrophy using unbiased computational analyses that allow for the identification of *Nkx2-5* dosage-sensitive pathogenic pathways in cardiac hypertrophy *in vivo*. In addition, a novel interacting co-regulator of Nkx2-5 and a functionally relevant downstream target gene were identified as involved in the pathogenesis of cardiac hypertrophy. Lastly, the present study demonstrates the utility of our computational approach in identifying novel mechanisms of differential gene regulation in the pathogenesis of complex diseases.

RESULTS

Nkx2-5$^{+/-}$ mice demonstrate exaggerated cardiac hypertrophy in response to pathological stimuli

Nkx2-5$^{+/-}$ mice express half-normal levels of *Nkx2-5* mRNA and protein [81,83]. The reduction in *Nkx2-5* dosage increases the risk of congenital heart defects (CHDs) [31] during fetal development. *Nkx2-5*$^{+/-}$ neonates without CHDs, however, grow into

healthy adults. Under basal conditions, adult *Nkx2-5*$^{+/-}$ hearts were indistinguishable from wild-type hearts in cardiac mass relative to body weight ratio (HW/BW, 4.3±0.1 mg/g in *Nkx2-5* WT/sham vs. 4.4±0.1 mg/g in *Nkx2-5*$^{+/-}$/sham; P = 0.25; Fig. 3.1a). The phenotype associated *Nkx2-5* haploinsufficiency became apparent only when the animals were subjected to adverse hypertrophic stimuli. Transverse aortic constriction (TAC) [27,84] performed on 8-10 week old wild-type mice resulted in an increase in HW/BW ratio compared with sham-operated littermates 10 days after surgery (4.3±0.1 mg/g in *Nkx2-5* WT/sham vs. 5.5±0.2 mg/g in *Nkx2-5* WT/TAC; P = 0.0004; Fig. 3.1a). After undergoing the same procedure, *Nkx2-5*$^{+/-}$ mice developed a 73% greater increase in HW/BW ratio compared with wild-type mice (28±5% increase in *Nkx2-5* WT/TAC vs. 49±6% increase in *Nkx2-5*$^{+/-}$/TAC; P = 0.01; Fig. 3.1a), suggesting that the development of appropriate cardiac hypertrophy in response to surgical perturbation is an *Nkx2-5* dosage-sensitive phenomenon.

A similar *Nkx2-5* haploinsufficiency phenotype was observed in the *calcineurin* overexpression (*CnA*-Tg)[21] model of cardiac hypertrophy. In five-week-old mice expressing normal cardiac levels of *Nkx2-5*, transgenic expression of *calcineurin* resulted in a doubling of heart weight in comparison to their non-transgenic (*CnA*-Ntg) littermates (HW/BW ratio 4.3±0.1 mg/g in *Nkx2-5* WT/*CnA*-Ntg vs. 9.0±0.2 mg/g in *Nkx2-5* WT/*CnA*-Tg; P =0.0004; Figs. 3.1b and 3.1c, Table 3.1). In *Nkx2-5*$^{+/-}$ mice, *calcineurin* overexpression tripled the heart weight relative to the non-transgenics (HW/BW ratio 4.3±0.1 in *Nkx2-5*$^{+/-}$/*CnA*-Ntg vs. 13.3±0.5 in *Nkx2-5*$^{+/-}$/*CnA*-Tg; P = 0.0001; Figs. 3.1b and 3.1c, Table 3.1). The hypertrophic response in *Nkx2-5*$^{+/-}$/*CnA*-Tg mice was therefore twice that of *Nkx2-5* WT/*CnA*-Tg mice (110±5% increase in *Nkx2-5* WT/*CnA*-Tg vs. 210±12% increase in *Nkx2-5*$^{+/-}$/*CnA*-Tg; P = 0.0001; Figs. 3.1b and 3.1c, Table 3.1). An increase in lung

weight to body weight ratio, indicative of pulmonary edema, was observed in *Nkx2-5*$^{+/-}$/*CnA*-Tg animals in association with the enhanced hypertrophic response (LW/BW ratio 6.0±0.1 mg/g in *Nkx2-5* WT/*CnA*-Tg vs. 6.9±0.1 mg/g in *Nkx2-5*$^{+/-}$/*CnA*-Tg; P =0.0001; Table 3.1). Similar to the TAC model, there was no difference in the heart weight to body weight ratio between *Nkx2-5*$^{+/-}$ and wild-type hearts under baseline conditions (4.3±0.1 mg/g in *Nkx2-5* WT vs. 4.3±0.1 mg/g in *Nkx2-5*$^{+/-}$; P =0.95 Figs. 3.1b and 3.1c, Table 3.1). Additional physiological parameters of *Nkx2-5*$^{+/-}$ mice under baseline and hypertrophic conditions are listed in Table 3.1. The replication of this *Nkx2-5* haploinsufficiency phenotype in disparate cardiac hypertrophy models suggests that *Nkx2-5*-dependent modification of the hypertrophic response is independent of the types of hypertrophic stimuli.

The aggravated hypertrophic response to *calcineurin* overexpression in *Nkx2-5*$^{+/-}$ animals compared with wild-type littermates can partially be attributed to the increases in individual cardiomyocyte size. Under non-hypertrophic conditions, the cross-sectional area of single cardiomyocytes measured from short-axis heart sections did not vary between *Nkx2-5* WT and *Nkx2-5*$^{+/-}$ hearts (43.8±1.6 μm^2 vs. 42.3±1.2 μm^2, respectively; P = 0.46; Fig. 3.1d). In the *calcineurin* overexpression model of cardiac hypertrophy, the cardiomyocyte cross-sectional area was increased by 57±7% in *Nkx2-5* WT/*CnA*-Tg hearts compared with Nkx2-5 WT/CnA-Ntg hearts (P < 0.0001; Fig. 3.1d), whereas the cardiomyocyte cross-sectional area in *Nkx2-5*$^{+/-}$/*CnA*-Tg hearts was >1.5 times greater than that in *Nkx2-5*$^{+/-}$/*CnA*-Ntg hearts (P < 0.0001; Fig. 3.1d). This translates to roughly three times more growth in cross-sectional area of individual cardiomyocytes as a result of *calcineurin* overexpression in *Nkx2-5*$^{+/-}$ hearts in comparison with *Nkx2-5* WT hearts. The enhanced cardiomyocyte hypertrophy observed in *Nkx2-5*$^{+/-}$ animals subjected to

hypertrophic stimuli was further corroborated by whole cell membrane capacitance recordings, which serve as a proxy measure of cardiomyocyte volume [85]. Cardiomyocytes isolated from *Nkx2-5* WT/*CnA*-Tg hearts exhibited 39±10% more capacitance than cardiomyocytes from *Nkx2-5* WT/CnA-Ntg hearts (132.6±7.5 pF vs. 184.9±7.1 pF, respectively; $P < 0.0001$; Fig. 3.1e). In *Nkx2-5*$^{+/-}$/*CnA*-Tg hearts, the capacitance of cardiomyocytes was 60±10% greater than that of cells from *Nkx2-5*$^{+/-}$/*CnA*-Ntg hearts (130.8±6.2 pF vs. 209.6±9.3 pF, respectively; $P < 0.0001$; Fig. 3.1e). *Nkx2-5*$^{+/-}$/*CnA*-Tg cardiomyocytes thus hypertrophied 54% more than cells from *Nkx2-5* WT/*CnA*-Tg hearts. The supernormal hypertrophic response in *Nkx2-5*$^{+/-}$ mice was additionally associated with impaired survival. Age of death as a result of *calcineurin* overexpression [86] in *Nkx2-5*$^{+/-}$/*CnA*-Tg mice was significantly younger than that in *Nkx2-5* WT/CnA-Tg littermates (196±8 days in *CnA*-Tg vs. 85±16 days in *Nkx2-5*$^{+/-}$/*CnA*-Tg; $P = 0.01$; Fig. 3.1f).

A subset of marker genes of cardiac hypertrophy was differentially regulated as a function of *Nkx2-5* dosage under hypertrophic conditions. Quantitative RT-PCR analysis revealed an enhanced transition from *a-myosin heavy chain* (*Myh6*) to *β-myosin heavy chain* (*Myh7*) in *Nkx2-5*$^{+/-}$ hearts, where *Myh7* expression level was elevated to twice the level seen in *Nkx2-5* WT hearts in response to *calcineurin* overexpression ($P = 0.01$; Fig. 3.1g, bottom panels). In contrast, up-regulation of both *ANF* and *BNP* mRNA levels were not significantly altered in *Nkx2-5*$^{+/-}$ hearts in comparison to wild-type hearts under hypertrophic conditions ($P = 0.44$ and 0.16, respectively; Fig. 3.1g, top panels). Notably, the level of *ANF* mRNA, but not *BNP* mRNA, was decreased in *Nkx2-5*$^{+/-}$ hearts in comparison to wild-type hearts under baseline conditions ($P = 0.01$ and 0.08, respectively), suggesting that *Nkx2-5* dosage sensitivity, in addition to being context-dependent, manifests differently for individual

genes. In a given cellular environment, such as that in the *CnA*-Tg heart, *Nkx2-5* haploinsufficiency only affects a subset of the molecular pathways involved in the hypertrophic response.

Transcriptional profiling captures Nkx2-5 dosage-sensitive transcripts in cardiac hypertrophy

A comprehensive survey of differential gene expression in *Nkx2-5*$^{+/-}$ and wild-type mice under basal and hypertrophic conditions would provide a foundation for identifying gene regulatory pathways that are sensitive to *Nkx2-5* dosage in the context of cardiac hypertrophy. To this end, global gene expression profiling was carried out in the ventricular myocardium of wild-type, *CnA*-Tg, *Nkx2-5*$^{+/-}$, and *Nkx2-5*$^{+/-}$/*CnA*-Tg mice using oligonucleotide microarrays. In addition, a comparable dataset from a TAC experiment [87] obtained through the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) was included in the ensuing computational analyses to minimize the likelihood of discovering model-specific artifacts. Supplementary Table S3.1 summarizes the results from the microarray study. The complete dataset is available for download through the Gene Expression Omnibus via accession number (pending submission). Of the 12,422 total probe sets represented on the microarray, 1,885 were significantly different in hearts subjected to TAC in comparison to sham operated hearts at an ANOVA p-value threshold of 0.05 (Fig. 3.2a). Of these probe sets, 727 were up-regulated and 1,158 were down-regulated. In the *calcineurin* overexpression model, significant differences in expression were detected between *CnA*-Tg mice and the non-transgenic littermates in 1,216 probe sets. Among these probe sets, 501 were up-regulated and 715 were down-regulated. Merging the two datasets produced 342 probe sets with significant differential expression in both the TAC and *calcineurin* overexpression models. 126 probe sets were up-regulated in both TAC- and *calcineurin* overexpression-induced cardiac

hypertrophy, including marker genes *angiotensin I converting enzyme 1 (Ace)* and *BNP*. Similarly, 162 probe sets were down-regulated in both models, including cardiac-specific homeodomain-only transcription regulator *Hopx*. The Gene Ontology (GO) biological process terms over-represented in the merged list of probe sets with significant differential expression in both models of cardiac hypertrophy are listed in Table 3.2.

The comparison between *Nkx2-5*$^{+/-}$ and wild-type hearts under baseline conditions revealed significant differences in the expression of 1,219 probe sets. Within this group of probe sets, 422 probe sets indicated up-regulation and 797 probe sets indicated down-regulation. While *Nkx2-5*$^{+/-}$ adults had no discernable phenotype differences from wild-type adults under baseline conditions, GO analysis applied to the expression profiles of non-hypertrophic *Nkx2-5*$^{+/-}$ and wild-type hearts nevertheless identified biological process terms associated with the differences in gene expression at a false discovery rate of 5%, including "protein polymerization" and "protein biosynthesis". The complete listing of GO terms associated with differential gene expression in *Nkx2-5*$^{+/-}$ hearts under baseline conditions is found in Table 3.2. Under hypertrophic conditions, comparisons among transcription profiles from *Nkx2-5* WT/*CnA*-Tg and *Nkx2-5*$^{+/-}$/*CnA*-Tg hearts revealed only 364 probe sets to be differentially expressed in a significant manner. Exactly one half of the probe sets in this grouping were up-regulated and the other half were down-regulated. The comparatively low number of differentially regulated genes suggests that the hypertrophic response as a result of *calcineurin* overexpression is not fundamentally different between *Nkx2-5*$^{+/-}$ and *Nkx2-5* WT mice. GO analysis of these probe sets points to an enrichment of several biological processes related to heart function, but none reached statistical significance at a 5% false discovery rate. When the 364 probe

sets from this comparison were merged with the 342 probe sets differentially regulated in both TAC and *CnA*-Tg models, 33 probe sets were found at the intersection. These represented genes differentially expressed in cardiac hypertrophy whose expression levels were further modified as a function of *Nkx2-5* dosage under hypertrophic conditions, presumably under the regulation of transcription regulatory pathways where Nkx2-5 has a dosage-dependent effect in the context of cardiac hypertrophy.

Computational analysis suggests Nkx2-5 exerts influence on the Egr1 pathway

The results of the transcription profiling experiments suggest that a reduced *Nkx2-5* level affects a limited subset of genes involved in the hypertrophic response. However, the transcriptional pathways regulating this subset of genes may regulate other target genes, independent of *Nkx2-5* dosage, in the pathogenesis of cardiac hypertrophy. We therefore devised a computational strategy to screen for transcription factors driving such regulatory pathways, using the entire complement of gene expression changes involved in the hypertrophic process and *Nkx2-5* haploinsufficiency (Fig. 3.2a). In this analysis, 394 experimentally determined mammalian transcription factor binding site (TFBS) profiles from the TRANSFAC database [88] were mapped to the proximal promoter sequences (from -2000bps to +500bps relative to annotated transcription start site) of each gene. The correlation between each TFBS's presence in the proximal promoter with the differential gene expression in a given mouse model was then evaluated using a ranksum test. TFBS signatures associated with gene expression changes both in normal transition to cardiac hypertrophy (in TAC vs. sham and *CnA*-Tg vs. *CnA*-Ntg comparisons) and in the escalation of hypertrophic response as a result of reduced *Nkx2-5* dosage (i.e. *Nkx2-5*$^{+/-}$/*CnA*-Tg vs. *Nkx2-5* WT/*CnA*-Tg comparisons) are indicative of *Nkx2-5*

dosage-sensitive transcriptional regulatory pathways active in cardiac hypertrophy. The comparison between *Nkx2-5*$^{+/-}$ and wild-type hearts was analyzed separately to identify Nkx2-5 dosage-sensitive pathways under baseline conditions.

The TFBS screening results are shown in Supplementary Table S3.2, and summarized in Figure 3.2a. In the comparison between TAC and sham-operated hearts, 45 unique human and/or mouse TFBS profiles ($P < 0.05$) associated with differential gene expression (regardless whether up- or down-regulation) were identified from a total of 394 TFBS candidates. Similarly, 64 unique TFBS profiles were found to be associated with expression changes between the *CnA*-Tg hearts and the *CnA*-Ntg hearts. Neither of the two profiles representing Nkx2-5 TFBS was found in these two lists, suggesting that direct Nkx2-5 binding to the proximal promoters may not be a dominant mechanism in these models of cardiac hypertrophy. A further set of 50 unique TFBS profiles was found to be associated with differential gene expression in the comparison between *Nkx2-5* WT/*CnA*-Tg and *Nkx2-5*$^{+/-}$/*CnA*-Tg hearts, including one of the TFBS representing Nkx2-5 (M00240, $P = 0.02$). The above TFBS lists were merged to reveal three TFBS profiles corresponding to androgen receptor (Ar), BTB and CNC homolog 2 (Bach2), and early growth response 1 (Egr1) in the intersection of all three sets (Fig. 3.2a). The presence of these binding signatures in the proximal promoters is associated with differential gene expression in both TAC and *calcineurin* overexpression models and with the *Nkx2-5* dosage-dependent modification of gene expression in *Nkx2-5*$^{+/-}$/*CnA*-Tg mice (Fig. 3.2b). Of these transcription factors, Egr1 was singled out for further characterization in the present study because of its well-defined role in the pathogenesis of cardiac hypertrophy [89].

A separate analysis comparing *Nkx2-5*$^{+/-}$ and *Nkx2-5* WT hearts identified 74 unique TFBS profiles that are over-represented in the proximal promoters of differentially regulated genes in Nkx2-5 haploinsufficiency under baseline conditions (Fig. 3.2a and Supplementary Table S3.2). One of the two profiles representing Nkx2-5 was present in this list (M00241, P = 0.05), suggesting that direct Nkx2-5 binding to the promoters may be necessary or sufficient in some of the *Nkx2-5* dosage sensitive gene targets. Other members of this list include important transcription regulators in the postnatal heart, such as CREB, Hif1, Mef2, NFκb, PPAR, and SRF. A reduced Nkx2-5 level may affect transcriptional regulation in the target genes via a modulatory effect on these transcription co-regulators. Notably, Egr1 was not found in this set. Furthermore, only four unique TFBS profiles were found in the intersection of the *Nkx2-5*$^{+/-}$ sets under baseline or hypertrophic conditions (cumulative P for $X \leq 4 = 0.02$, Fig. 3.2a). Taken together, the results from the *Nkx2-5*$^{+/-}$ TFBS screens suggest *Nkx2-5* haploinsufficiency affects distinct transcriptional pathways in the heart depending on the pathological context.

The TFBS screen in the hypertrophic hearts led to a hypothetical model where Nkx2-5 exerts a dosage-dependent effect on Egr1-dependent regulation of transcription in the pathogenesis of cardiac hypertrophy. *Egr1* is an immediate-early transcription factor activated in cardiac hypertrophy [90,91] and a key mediator of the hypertrophic process. Saadane et al. demonstrated that homozygous deletion of *Egr1* attenuates hypertrophic response to TAC [89]. Similarly, animals overexpressing Egr1-specific suppressor *NGFI-A binding protein 1* (*Nab1*) responded to TAC with markedly reduced hypertrophy [92]. In the *calcineurin* overexpression model, the expression level of *Egr1* was elevated under hypertrophic conditions (P = 0.02; Fig. 3.2c). The baseline expression and the up-regulation of *Egr1* in cardiac hypertrophy

was unaffected by *Nkx2-5* gene dosage (P = 0.51 under baseline conditions and P = 0.37 under hypertrophic conditions; Fig. 3.2c). In comparison, the *Nab1* expression level remained constant in cardiac hypertrophy and was neither a function of *Nkx2-5* dosage under baseline nor hypertrophic conditions (Fig. 3.2c). *NGFI-A binding protein 2 (Nab2)* is a *Nab1* paralog with overlapping functional roles [93]. *Nab2* transcription is directly regulated by Egr1 [47]. As measured with both microarrays and quantitative RT-PCR, *Nab2* expression was up-regulated in cardiac hypertrophy and was further elevated in *Nkx2-5*$^{+/-}$/*CnA*-Tg hearts (*CnA*-Ntg vs. *CnA*-Tg, P = 0.02; *Nkx2-5* WT/*CnA*-Tg vs. *Nkx2-5*$^{+/-}$/*CnA*-Tg, P = 0.03; Fig. 3.2c). *Nab2* expression level, however, was not responsive to *Nkx2-5* dosage under baseline conditions (P = 0.86; Fig. 3.2c). Lastly, the *Nkx2-5* level was halved, as expected, in *Nkx2-5*$^{+/-}$ hearts under both baseline and hypertrophic conditions (P = 0.02 and 0.05, respectively; Fig. 3.2c). *Nkx2-5* was not differentially regulated in response to *calcineurin* overexpression regardless of *Nkx2-5* dosage.

Nkx2-5 physically interacts with Egr1 and Nab1

To establish the role for Nkx2-5 as a co-regulator of Egr1-dependent transcriptional activation, an *in vitro* binding assay was carried out to assess potential protein-protein interactions between Nkx2-5 and Egr1 (Fig. 3.3a). ^{35}S-labeled Egr1 and Egr1 suppressor Nab1 were both shown to bind to Nkx2-5 fused to maltose binding protein (MBP-Nkx2-5) immobilized on amylose beads. Known interactions between Nab1 and Egr1 [94], and between Gata4 and Nkx2-5 [95], were demonstrated as positive controls. ^{35}S-labeled Gata-4 did not bind to MBP-Egr1, indicating that specificity was maintained in the assay. Additionally, the Nkx2-5 protein domain required for the physical interaction between Nkx2-5 and Egr1 was determined with an *in vitro* binding assay between MBP-Egr1 and ^{35}S-labeled Nkx2-5 mutants with N-

and C-terminal truncations for known functional domains. As shown in Figure 3.3b, only those Nkx2-5 fragments containing the homeodomain were able to bind to MBP-Egr1 *in vitro*. The homeodomain had previously been shown to be necessary for the protein-protein interaction between Nkx2-5 and Gata-4 [95], despite being primarily known for its DNA binding activity. As a control, the same set of ^{35}S-labeled Nkx2-5 deletion mutants was tested against MBP-Nkx2-5 to recapitulate results from previous reports that Nkx2-5 homodimerizes via an interaction between homeodomains [95,96]. The NK2SD domain adjacent to the homeodomain in the C-terminal direction contributes to the Egr1-Nkx2-5 interaction and Nkx2-5 homodimerization. Nkx2-5 fragment B, which retained the homeodomain but not the NK2SD domain, was able to bind to both MBP-Egr1 and MBP-Nkx2-5, albeit with reduced affinity.

Nkx2-5, along with Nab1, modulate Egr1-dependent transcriptional activation

Nab2 expression is regulated by Egr1 [47] and is sensitive to *Nkx2-5* dosage in the context of *calcineurin* overexpression-induced cardiac hypertrophy (Fig. 3.2c). To ascertain the function of Egr1-Nkx2-5 interaction, a transactivation assay was carried out in C2C12 myoblasts using a luciferase reporter construct driven by the murine Nab2 proximal promoter (Fig. 3.3c). The introduction of exogenous Egr1 alone activated the luciferase reporter as expected ($P < 0.0001$). The addition of Nkx2-5 or Nab1 had no effect on the baseline activity of the *Nab2* promoter, but both were able to attenuate Egr1-activated reporter activity ($P < 0.0001$ for both). Egr1-dependent activation of the *Nab2* promoter was maximally suppressed when Nkx2-5 and Nab1 were added to the reaction in combination ($P < 0.0001$).

Given that the paralogs Nab2 and Nab1 share structural and functional features and may be partially interchangeable, it is conceivable that Nab2 may itself exert some effect on Egr1-dependent transcriptional regulation. A follow-up

transactivation experiment showed that the addition of Nab2 diminished the collective repressive effect by Nkx2-5 and Nab1 on the Egr1-dependent activation of the *Nab2* promoter (Fig. 3.3d). This modulatory effect of Nab2 on Egr1-dependent transcriptional regulation was dependent on Nab2 concentration ($R = 0.88$; $P = 0.0002$). In contrast, the addition of GFP in increasing concentrations had no effect on the Egr1-dependent activation of the *Nab2* promoter ($R = -0.22$; $P = 0.59$). These findings suggest that the transcriptional activation of *Nab2* may constitute a feed forward feature in Egr1-dependent transcriptional regulation in the pathogenesis of cardiac hypertrophy.

Heterozygous deletion of *Nab2* reduced the severity of cardiac hypertrophy in *CnA*-Tg mice

The results from the *in vitro* transactivation experiments predicted that a decrease in *Nab2* expression may impede Egr1-dependent activation of target gene expression and thus curtail the development of cardiac hypertrophy. We tested this hypothesis genetically by measuring the extent of cardiac hypertrophy in 35-day old wild-type, $Nab1^{+/-}$ and $Nab2^{+/-}$ mice overexpressing *calcineurin*. Under baseline conditions, the heart weight to body weight ratio in both $Nab1^{+/-}$ and $Nab2^{+/-}$ mice was not significantly different from with the ratio in their wild-type counterparts (4.8±0.1 mg/g in wild-type vs. 4.7±0.1 mg/g in $Nab1^{+/-}$; $P = 0.72$, and vs. 4.7±0.1 mg/g in $Nab2^{+/-}$; $P = 0.66$; Fig. 3.4a). In *calcineurin* overexpression-induced cardiac hypertrophy, $Nab2^{+/-}$ mice had a 10.4±3% lower HW/BW ratio in comparison to their wild-type littermates (11.4±0.3 mg/g in *Nab2*-WT/*CnA*-Tg vs. 10.2±0.2 mg/g in $Nab2^{+/-}$/*CnA*-Tg; $P = 0.05$; Fig. 3.4a). In contrast, the HW/BW ratio in $Nab1^{+/-}$ mice was not significantly different from that in the wild-type mice under hypertrophic conditions (11.4±0.3 mg/g in *Nab1*-WT/*CnA*-Tg vs. 11.2±0.3 mg/g in $Nab1^{+/-}$/*CnA*-Tg; $P = 0.62$; Fig. 3.4a). In contrast to *Nkx2-5*, *Nab1* and *Nab2*

haploinsufficiency did not lead to an increased lung weight to body weight ratio under hypertrophic conditions (Fig. 3.4a). Reduced *Nab2* dosage also affected mean survival time in animals with cardiac hypertrophy. When subjected to *calcineurin* overexpression-induced hypertrophic stress, $Nab2^{+/-}$ mice demonstrated a prolonged lifespan relative to their wild-type and $Nab1^{+/-}$ littermates (149±17 days in $Nab2^{+/-}$/*CnA*-Tg vs. 96±12 days in *CnA*-Tg; P = 0.02, and vs. 124±25 days in $Nab1^{+/-}$/*CnA*-Tg; P = 0.03; Fig. 3.4b).

The expression pattern of the cardiac hypertrophy marker genes *ANF* and *Myh7*, but not *BNP*, reflected the attenuated hypertrophic response to *calcineurin* overexpression in $Nab2^{+/-}$ mice. The up-regulation associated with pathological cardiac hypertrophy of *ANF* and *Myh7* was blunted in $Nab2^{+/-}$ hearts in comparison to wild-type hearts in a quantitative RT-PCR analysis (P = 0.01 and 0.05; Fig. 3.4c). mRNA expression of transcriptional regulators was also differentially regulated to varying degrees in *Nab* animals. The elevated *Egr1* mRNA level under hypertrophic conditions did not vary significantly among wild-type, $Nab1^{+/-}$, and $Nab2^{+/-}$ hearts (Fig. 3.4d). Similarly, *Nkx2-5 and Nab1* expression levels were not responsive to a reduced dosage of Nab2 or the hypertrophic status of the hearts (Fig. 3.4d). *Nab2* expression, however, was up-regulated in cardiac hypertrophy except in $Nab2^{+/-}$ hearts (*CnA*-Ntg vs. *CnA*-Tg, P = 0.02; $Nab1^{+/-}$ vs. $Nab1^{+/-}$/*CnA*-Tg, P = 0.01; $Nab2^{+/-}$ vs. $Nab2^{+/-}$/*CnA*-Tg, P = 0.32; Fig. 3.4d). Notably, *Nab2* mRNA level was elevated in $Nab1^{+/-}$ hearts, irrespective of the hearts' state of hypertrophy (*Nab1* WT/*CnA*-Ntg vs. $Nab1^{+/-}$/*CnA*-Ntg, P = 0.01; *Nab1* WT/*CnA*-Tg vs. $Nab1^{+/-}$/*CnA*-Tg, P = 0.05), suggesting that Nab1-dependent suppression of *Nab2* expression is independent of cardiac hypertrophy. Together, these results point to a context-dependent network of regulation arising from an intricate interplay among Egr1, Nab1, Nab2, and Nkx2-5.

DISCUSSION

Nkx2-5 has been extensively studied in fetal heart development but remains relatively unexplored in adults. In this study, we hypothesized that postnatal *Nkx2-5* expression may modulate the pathogenesis of adult heart disease, so we set out to characterize *Nkx2-5*$^{+/-}$ mice in the context of cardiac hypertrophy. *Nkx2-5*$^{+/-}$ mice express half the normal level of *Nkx2-5* mRNA and protein, which presumably approximates naturally-occurring variations in the normal physiological range of *Nkx2-5* levels more closely than either knockout or overexpression models [97]. In the absence of hypertrophic stress, adult *Nkx2-5*$^{+/-}$ mice with no CHDs lack a readily detectable morphological or physiological phenotype. When challenged with pathologic surgical or genetic stimuli, however, *Nkx2-5*$^{+/-}$ mice exhibited markedly enhanced cardiac hypertrophy and poorer survival outcomes than their wild-types littermates. A set of unbiased computational screens was applied to identify the mechanisms underlying the *Nkx2-5* dosage dependence of hypertrophic responses in the heart. These analyses combined orthogonal datasets of disparate experimental origins to generate a mechanistic hypothesis that *Nkx2-5* influences the pathogenesis of cardiac hypertrophy in part by modulating the transcription program regulated by Egr1, a known mediator of cardiac hypertrophy. Experimental validations in this study subsequently confirmed that Nkx2-5 physically interacts with Egr1 and its specific repressor Nab1 to suppress Egr1-dependent transcriptional activation in cardiac hypertrophy. Whether Nkx2-5-dependent repression of Egr1-dependent transcription requires physical interaction between Nkx2-5 and DNA in the target promoter remains an open question. Computational dynamics modeling and further

experimental investigations are required to reveal the full mechanistic details of the Nkx2-5-Egr1-Nab1 interaction.

Nab2, a *Nab1* paralog, was identified in our screen as a prototypical *Nkx2-5* dosage-sensitive target gene under Egr1-dependent transcriptional regulation in the context of pathological cardiac hypertrophy. In our mouse models, the mRNA expression level of *Nab2* corresponds to the severity of hypertrophic response. A causal relationship between elevated *Nab2* expression and increased cardiac hypertrophy is supported by a transactivation experiment in which the addition of Nab2 negated the suppression of Egr1 activity by Nab1 and Nkx2-5 in a concentration-dependent manner. The effect of Nab2 on the pathogenesis of cardiac hypertrophy was confirmed genetically in the $Nab2^{+/-}$/*CnA*-Tg mice, in which reduced *Nab2* dosage attenuated the hypertrophic response to *calcineurin* overexpression and prolonged survival. Egr1-dependent expression of *Nab2* may constitute a novel feed forward mechanism in the pathogenesis of cardiac hypertrophy. However, Nab2 has also previously been described as a suppressor of Egr1-dependent transcription [98]. The role for Nab2 may therefore be context-dependent, and requires further characterization. The novel findings from this study are summarized in Figure 3.5.

In the last two decades, over-expression and knockout models have been established as invaluable tools to study gene function. The expression levels of genes under study, from nearly zero in null models to up to several thousand times the wild-type level in overexpression models, departs drastically from what is physiologically plausible in complex biological processes with multiple layers of amplification and feedback regulation. The mechanistic insights derived from the dramatic phenotype in these studies are often qualitative in nature. For instance, previous studies of *Nkx2-5* function in the postnatal heart have linked large deviations

from the physiological range of *Nkx2-5* dosage to the development of cardiomyopathy [34,79,82]. The mechanisms associated with these dramatic phenotypes, however, may not be directly relevant in the understanding of similar pathologies in human patients. Modern genomic methods, with ever-greater capacities to generate larger and more accurate datasets, greatly enhance our ability to identify mechanisms underlying subtle and previously undetectable phenotypic observations. The modest scale of differential gene expression involved in governing such regulatory mechanisms belies their true biological significance. As an example, our study highlights the importance of precise gene dosage regulation in the upstream regulators *Egr1* and *Nkx2-5* in the context of a major disease process. In general, a haploinsufficient animal model that is indistinguishable from wild-type littermates under normal conditions but that responds differently under pathological conditions may provide a closer approximation of real world health and disease than either null models or overexpression models can deliver.

Lastly, we demonstrated the utility of computational screens as a tool to rapidly generate testable hypotheses in the laboratory. In this study, computational analyses were carried out using data generated internally as well as data freely available from public depositories. The analyses were designed to target a specific class of hypothetical mechanisms underlying the phenotypic observations manifested in the *Nkx2-5*$^{+/-}$ mice. Validation experiments were then performed to inform the subsequent computational analyses. This iterative and hybrid approach for solving biological problems may substantially accelerate discoveries and reduce research costs; we were able to identify novel molecular interactions among transcriptional regulators Egr1, Nab1, and Nkx2-5, and a functionally relevant downstream target gene, *Nab2,* despite the limitations inherent in our computational screen, which was designed to reduce

the complexity of the analyses. These limitations include: focusing solely on transcriptionally regulated mechanisms, using stringent criteria to exclude additional candidate transcriptional regulators and target genes in the TFBS screen, and limiting the analyses to relatively small regions in the proximal promoters. Given the ever-increasing volume and quality of data deposited in the public domain, we expect computational screens to become a routine and ubiquitous part of biomedical research. Further developments of this computational method are necessary in order to expand its capabilities and enhance its usability to capitalize on the deluge of data from modern biomedical studies.

METHODS

Mouse husbandry and genotyping

Nkx2-5 knockout mice were generated by targeted replacement of the entire coding region with β-galactosidase, as has been previously described [80]. The *Nkx2-5* strain is maintained in a C57Bl/6 background after >6 backcrosses. *Calcineurin* transgenic mice were kindly provided by Dr. Jeff Molkentin, and were maintained by mating transgenic males to hybrid B6C3F1 females [21]. B6C3F1 mice were purchased from Charles River Laboratories. *Nab1* and *Nab2* mice [99] were a generous gift from Dr. Jeff Milbrandt (Washington University School of Medicine, St. Louis, MO). All protocols conformed with the protocols of the Association for the Assessment and Accreditation of Laboratory Animal Care and are approved by the animal studies committee at Washington University School of Medicine.

Transverse aortic constriction model of cardiac hypertrophy

Male mice, aged 8-10 weeks and weighing 22-25 g, were subjected to transverse aortic banding to induce cardiac hypertrophy, or to a sham operation.

Pentobarbital (70 mg/kg intraperitoneal) was given for general anesthesia. Each mouse was positioned supine, and the trachea orally intubated for mechanical ventilation. After subcutaneous injection of 1% lidocaine (0.1 ml) for local anesthesia, an incision was made into the left second intercostal space. The aortic arch was mobilized. The right innominate, left carotid and left subclavian arteries clearly visualized. A 7-0 silk suture was passed around the transverse aorta between the innominate and left carotid arteries and tied against a 27-gauge needle to create a stenotic band. The needle was removed and the chest closed by suturing each layer of muscle and skin. The mouse remained intubated until resumption of spontaneous respiration. The animals were allowed to recover under warming lights or on warming pads until they were fully awake. Sham-operated mice underwent the entire procedure, including dissection of the aortic arch, but not banding. The duration of the operation was 20-25 minutes.

Heart weight measurement

Subjects were sacrificed between 35-39 days of age. Measurement of the body weight, heart weight, lung weight, and tibia length were taken immediately following dissection.

Cell culture

The C2C12 pre-myocyte cell line was purchased from the ATCC (Manassas, Virginia) and maintained according to protocol supplied by the vendor (http://www.atcc.org/). C2C12 cells were transfected using Lipofectamine 2000 (Invitrogen) as per the manufacturer's instructions. Transfected cells were maintained for an additional 36 hours following transient transfection before harvest.

RNA extraction

RNA was isolated from the ventricles using Trizol reagent (Invitrogen)

following the manufacturer's protocol. Ventricular tissues stored in -80^{o}C were homogenized in Trizol using the PowerGen 125 homogenizer (Fisher Scientific). The precipitated RNA was resuspended in 175 μL of RNA lysis buffer from the SV Total RNA Isolation Kit (Promega, Madison, WI) and purified according to the manufacturer's protocol. cDNA was synthesized from 2.5 μg of total RNA using SuperScript VILO cDNA Synthesis Kit (Invitrogen) according to manufacturer's protocol.

Quantitative PCR

Quantitative PCR of cDNA was performed in Mx3005P QPCR System (Stratagene, Santa Clara, CA) in a 40μL reaction containing iQ SYBR Green Supermix (Bio-Rad) and PCR primers at a final concentration of 200 nM. Primer sequences were designed with the PerlPrimer [75] or obtained from PrimerBank [76]. Primers were purchased from Sigma Aldrich or Integrated DNA Technologies (Coralville, IA). The primer sequences are provided in the supplemental methods. Relative mRNA expression was determined using the ΔΔ-Ct method and normalized to *Actb*.

Expression profiling

Expression profiles of four *calcineurin* transgenic ventricles (*CnA*-Tg) and four non-transgenic ventricles (*CnA*-Ntg) from five-week-old males were generated using Affymetrix U74Av2 chips. Expression profiles were scaled with the MAS5.0 software (Affymetrix) and normalized by mean. Expression profiles for FVB mice subjected to either TAC or sham operations were obtained from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) [39] under series accession number GSC1621 [87]. Differential gene expression was detected using profile ANOVA in

Spotfire DecisionSite v9.1 (TIBCO, Palo Alto, CA). GO terms over-represented in differentially expressed genes were identified using Genemerge [74].

Computational analysis

An algorithm modified from Mootha et al. [100] was used to correlate the presences of mammalian TRANSFAC motifs [88] within the 2.5 kb of proximal promoters (-2000bps to +500bps relative to annotated transcriptional start sites) to differential gene expressions. Promoter sequences were extracted from mouse genome mm8 build based on Refgene annotations and assigned a *significance rank* derived from the p-values associated with differential expression in the genes they represent in a particular expression profile experiment. For each transcription factor binding site (TFBS) profile, the algorithm first identifies the best instance within each proximal promoter sequence as previously described [101-103] at every base position, both in the forward and reverse orientations. A dynamic score threshold is then used to make presence/absence calls for that TFBS profile within each promoter. The algorithm thus does not take into account multiple occurrences of weaker TFBSs in a given promoter. For each TFBS profile, the significance rank associated with presence and absence calls obtained at every step of bit-score thresholding is evaluated using ranksum statistics. P-values from the ranksum test were corrected for multiple hypotheses testing using false discovery rate calculations.

***In vitro* binding assay**

MBP-fusion proteins were expressed in BL21 DE3 competent cells in rich media containing 0.3mM IPTG. Fusion proteins were released by sonication and bound to amylose resins. ^{35}S-methionine-labeled Egr1, Nkx2-5, Nab1, and Gata-4 were expressed using T7-coupled reticulocyte lysate system (Promega). Radiolabeled proteins were incubated in binding buffer with fusion protein bound amylose resins

for 4 hours in 4°C and electrophoresed by SDS-PAGE. Binding between radiolabeled proteins and MBP-fusion proteins were detected with autoradiography.

Measurement of whole-cell membrane capacitance

Myocytes were isolated using procedures previously described [104]. Briefly, hearts were mounted on a Langendorf apparatus and perfused in a retrograde direction through the aorta with a buffer containing 0.8 mg/ml collagenase (type II, Worthington Biochemical Corp., Lakewood, NJ) [105]. Following the perfusion, the LV apex was dissected, mechanically dispersed and plated on laminin-coated coverslips and maintained in 5% CO_2. Whole-cell recordings were obtained from LV apex myocytes within 24 hours at room temperature. All voltage-clamp experiments were performed using an Axopatch 1B patch clamp amplifier (Molecular Devices, Sunnyvale, CA) and the pCLAMP9 software package (Molecular Devices). Data were filtered at 5 kHz before storage. Voltage-clamp data were compiled and analysed using Clampfit v.9.2 (Molecular Devices) and Microsoft Excel. Capacitative transients recorded in brief ±10 mV voltage steps from the holding potential (−70 mV) were integrated to yield the whole-cell membrane capacitance (Cm).

Measurement of Cardiomyocyte Cross-sectional Areas

Mice were anesthetized with a terminal dose (120 mg/kg) of sodium pentobarbital via intraperitoneal injection. The sedated animals were perfused with 10 mL of phosphate buffered saline containing 10 mg of heparin (Sigma), followed by 100 mL of 4% paraformaldehyde (PFA) in 0.1 M sodium phosphate buffer (pH 7.2 at 20 °C) in a 20 mL/minute drip using a peristaltic pump. The hearts were then excised and fixed in 4% PFA for an additional hour. After dehydration in 70% ethanol overnight, the fixed hearts were embedded, sectioned, and stained with Masson's trichrome. Microscopy images (20X objective lens) of the heart sections were

captured using Zeiss Axiovision software. For each heart, 60 cardiomyocytes oriented perpendicular to the section plane were highlighted using Adobe Photoshop. The cross-sectional area of each highlighted cells were then determined using NIH ImageJ.

Cloning and Plasmids

Dr. Jeff Milbrandt (Washington University School of Medicine, St. Louis, MO) generously provided the following clones: pCB6 HA-*Egr1*, pCITE *Egr1*, FCIV HA-*Nab1*, CMVneo His-*Nab1*, and pcDNA3 *Nab2*. Murine *Nkx2-5* in pcDNA3.1 and pMAL vectors were a gift from Dr. Seigo Izumo. SV40-renilla luciferase in pGL4 was obtained through Promega. *Egr1*, *Nkx2-5* and *Nkx2-5* fragments were cloned from the expression vectors into pMALc2 vector (NEB) for *in vitro* expression of maltose binding protein (MBP) fusion proteins. Primers used for cloning *Nkx2-5* fragments are listed in Supplementary Table S3.3. *Nab2* promoter from -2000 bps to +500 bps was cloned from genomic DNA of a C57Bl/6 donor by PCR using primers listed in Supplementary Table S3.3. PCR product was cloned into pCRII cloning vector using the PCRII TOPO TA Cloning Kit (Invitrogen). The *Nab2* promoter was subsequently sub-cloned into a pGL3 luciferase vector (Promega) at the unique XhoI and HindIII site. Clones in 'plus' orientation were screened by restriction digestions and sequence verified.

Transactivation Assay

C2C12 pre-myocytes cultured in 24 well plates were transfected with 0.4mg of Nab2 luciferase reporter construct and 0.01mg of SV40 *renilla* luciferase construct, with or without activators Egr1, Nkx2-5, and Nab1. The transactivation assay was carried out immediately following cell harvest using the Dual-Luciferase Reporter Assay System (Promega) following the manufacturer's instructions. *Nab2* luciferase

activity was measured in 10 seconds intervals in a Leader 50i luminometer (Gen-Probe, San Diego, CA) serially. *Nab2* luciferase activity measurements were normalized to SV40-driven *renilla* luciferase readings measured in the same reaction.

Statistics

Two-tailed student's t-tests were used to access the significance level in all pair-wise comparisons. The significance levels of Venn logic were calculated using cumulative probability assuming hypergeometric distribution. All errors were reported as ± s.e.m.

ACKNOWLEDGEMENTS

The authors wish to thank Katia Valkova for her technical assistance. IDC was supported in part by the Lucille P. Markey Pathway in Human Pathobiology. PYJ was supported in part as a Scholar of the Child Health Research Center of Excellence in Developmental Biology at Washington University School of Medicine (K12-HD001487). This work was supported in part by grants from the Charles H. Hood Foundation, Hartwell Foundation, American Heart Association, March of Dimes (1-FY07-453), NIH (HL105857) and the Washington University Nutrition Obesity Research Center (NIH P30DK056341).

FIGURES

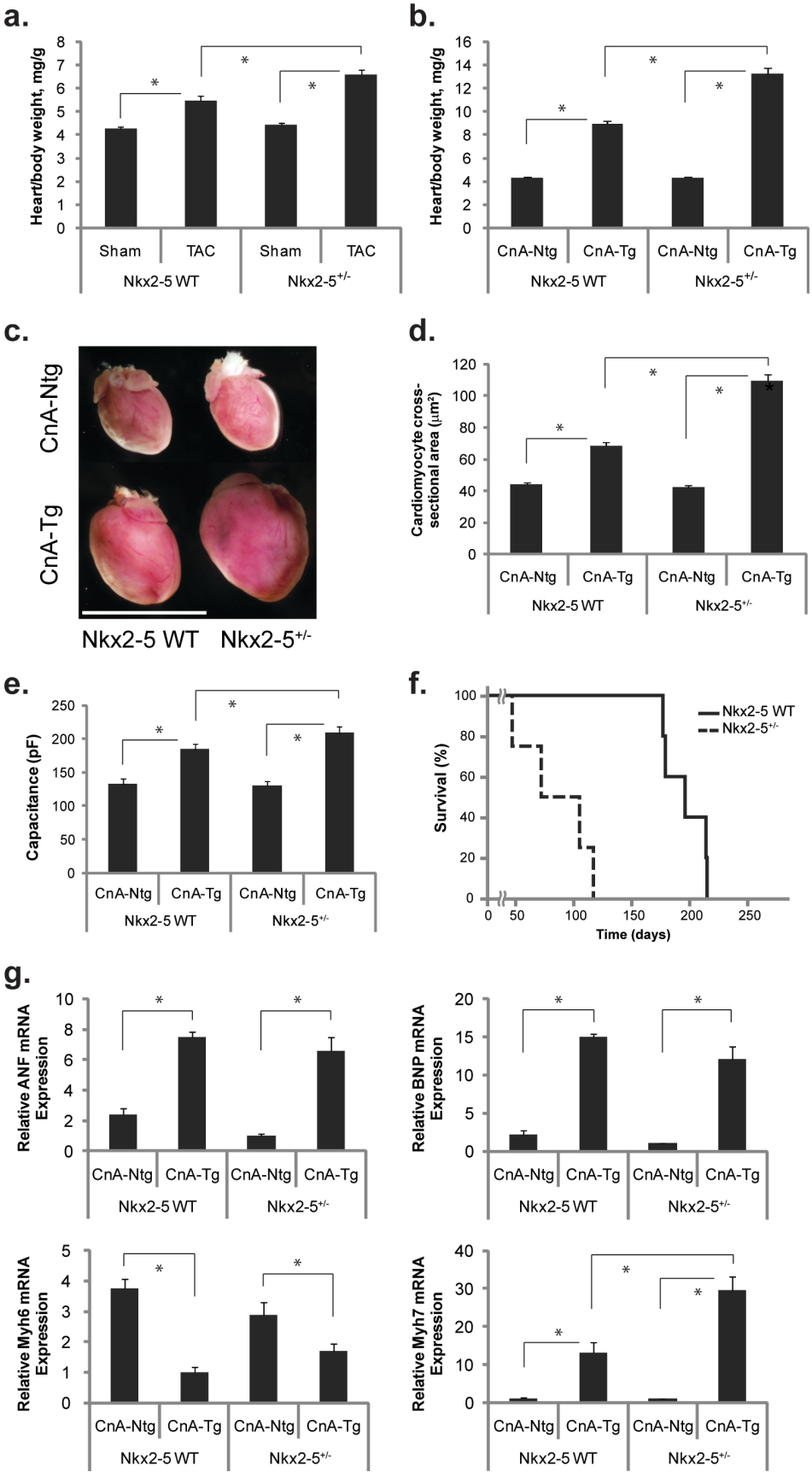

a.
Heart/body weight, mg/g
Sham
TAC
Sham
TAC
Nkx2-5 WT
Nkx2-5+/-
b.
Heart/body weight, mg/g
CnA-Ntg
CnA-Tg
CnA-Ntg
CnA-Tg
Nkx2-5 WT
Nkx2-5+/-
c.
CnA-Ntg
CnA-Tg
Nkx2-5 WT
Nkx2-5+/-
d.
Cardiomyocyte cross-sectional area (μm²)
CnA-Ntg
CnA-Tg
CnA-Ntg
CnA-Tg
Nkx2-5 WT
Nkx2-5+/-
e.
Capacitance (pF)
CnA-Ntg
CnA-Tg
CnA-Ntg
CnA-Tg
Nkx2-5 WT
Nkx2-5+/-
f.
Survival (%)
Nkx2-5 WT
Nkx2-5+/-
Time (days)
g.
Relative ANF mRNA Expression
Relative BNP mRNA Expression
Relative Myh6 mRNA Expression
Relative Myh7 mRNA Expression
CnA-Ntg
CnA-Tg
CnA-Ntg
CnA-Tg
Nkx2-5 WT
Nkx2-5+/-

Figure 3.1: *Nkx2-5* haploinsufficiency led to the exacerbated cardiac hypertrophy and poor prognosis in response to pathological hypertrophic stimuli. (a) Cardiac hypertrophy developed following transverse aortic constriction was enhanced In *Nkx2-5* haploinsufficient (*Nkx2-5*$^{+/-}$) mice in comparison to wild-type (WT) littermates (p=0.011). *Nkx2-5* haploinsufficiency has no appreciable phenotype in sham-operated mice. (b) The extent of cardiac hypertrophy developed following the overexpression of *calcineurin* transgene was roughly doubled in *Nkx2-5*$^{+/-}$/*CnA*-Tg hearts in comparison to *CnA*-Tg hearts (p<0.001). *Nkx2-5* WT and *Nkx2-5*$^{+/-}$ hearts showed no difference in terms of cardiac mass under baseline conditions. *n* = 6-18 per genotype group. (c) *Nkx2-5*$^{+/-}$/*CnA*-Tg hearts showed dramatic hypertrophy in comparison to wild-type hearts as a result of *calcineurin* overexpression at 35 days of age (bottom row). *Nkx2-5* WT and *Nkx2-5*$^{+/-}$ hearts were indistinguishable morphologically under baseline conditions (top row). The white scale bar indicates 10mm. (d) Cross-sectional area of ventricular cardiomyocytes was greater in the *Nkx2-5*$^{+/-}$/*CnA*-Tg hearts in comparison to *CnA*-Tg hearts. *Nkx2-5* dosage had no effect on the cross-sectional areas in non-transgenic (*CnA*-Ntg) cardiomyocytes. n = 60 per group. (e) Whole cell membrane capacitance was increased by 39% and 60%, respectively, in the *CnA*-Tg and *Nkx2-5*$^{+/-}$/*CnA*-Tg hearts in comparison to their non-transgenic counterparts (p=0.038). *Nkx2-5* dosage had no effect on the whole cell membrane capacitance values in non-transgenic (*CnA*-Ntg) cardiomyocytes. *n* = 60 per group. (f) *CnA*-Tg mice die prematurely due to heart failure (median = 196 days, solid tracing). The median survival time for *Nkx2-5*$^{+/-}$/*CnA*-Tg mice was reduced to 72 days for the male (*n* = 4, dark dashed tracing, p=0.005). (g) Quantitative RT-PCR result showing mRNA expression level for selected cardiac hypertrophy markers differentially regulated in *calcineurin* overexpression induced cardiac hypertrophy

and *Nkx2-5* haploinsufficiency ($n = 4$). For all panels, error bars indicate s.e.m. (*, $P < 0.05$).

	Nkx2-5 WT		*Nkx2-5*$^{+/-}$	
	CnA-Ntg	*CnA*-Tg	*CnA*-Ntg	*CnA*-Tg
HW/BW	4.27±0.08	8.95±0.22†	4.28±0.06	13.27±0.47*‡
LW/BW	6.15±0.06	5.95±0.14	6.26±0.11	6.88±0.06*‡
EDD	3.16±0.15	4.15±0.17†	3.12±0.10	4.21±0.14*
ESD	1.41±0.19	2.72±0.27†	1.67±0.15	2.71±0.21*
LVPW	0.65±0.03	0.82±0.02†	0.67±0.03	0.94±0.03*‡
FS	55.83±4.08	35.17±4.41†	46.89±3.45	36.17±3.39
HR	527.50±25.88	445.17±36.11	485.67±32.67	439.17±53.94
n	6	6	7	6

Table 3.1: Physiologic parameters in wild-type (WT) and *Nkx2-5*$^{+/-}$ mice under baseline (*CnA*-Ntg) or hypertrophic (*CnA*-Tg) conditions. HW/BW: heart weight to body weight ratio (mg/g); LW/BW: heart weight to body weight ratio (mg/g); EDD: end-diastolic diameter (mm); ESD: end-systolic diameter (mm); LVPW: left ventricular posterior wall thickness (mm); FS: fractional shortening (%); HR: heart rate (min^{-1}). †: $P < 0.05$ in the *Nkx2-5* WT/*CnA*-Ntg vs. *Nkx2-5* WT/*CnA*-Tg comparisons; ‡: < 0.05 in the *Nkx2-5* WT/*CnA*-Tg vs. *Nkx2-5*$^{+/-}$/*CnA*-Tg comparisons; *: < 0.05 in the *Nkx2-5*$^{+/-}$/*CnA*-Ntg vs. *Nkx2-5*$^{+/-}$/*CnA*-Tg comparisons.

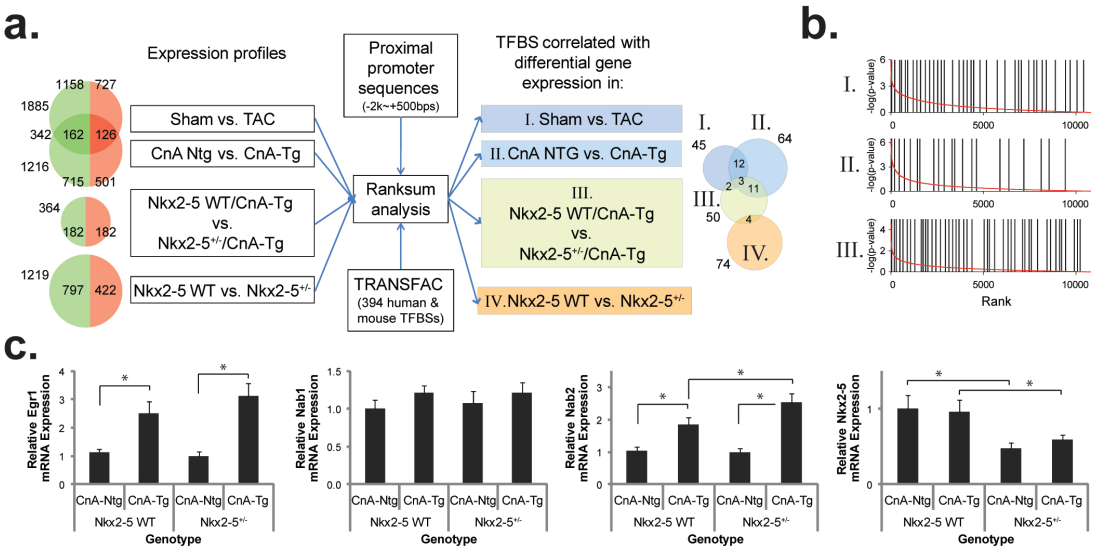


Figure 3.2: Egr1 was identified in a computational screen as an *Nkx2-5* dosage-sensitive transcriptional regulator in the pathogenesis of cardiac hypertrophy. (a) A schematic of our computational strategy and results. Left: the numbers of probe sets up- and down-regulated in each of the comparisons are given next to the Venn diagrams adjacent to the model names. Center: the proximal promoter sequences for every gene represented on the microarray were curated and ordered based on the significance level of their differential regulation in each comparison. A ranksum test was then used to determine the correlation between the presence of a particular transcription factor binding site (TFBS) in the proximal promoters and differential gene expression. Right: The numbers of TFBS whose presence in the proximal promoters was associated with differential gene regulation at a false discovery rate of 5% are given for each comparison in the Venn diagram. Egr1 was found in the intersection of all three TFBS lists. (b) The presence of Egr1 TFBS in the proximal promoters was associated with differential gene expression in all three comparisons (see Fig. 3.2a). The ranks of each gene, or probe set, ordered by the significance level of its differential expression in a given comparison, are represented on the horizontal

axis, with the most significant towards the left of the graph. The red tracings reflect the negative log of each gene's ANOVA p-value in a given comparison. Each black hatch represents five successive Egr1 TFBS positive proximal promoters and is positioned at the centroid of the five promoters on x-axis. The p-values of the correlations are 0.001 (ranked 11^{th}), 0.0001 (ranked 1^{st}), and 0.002, (ranked 16^{th}) in TAC, *CnA*-Tg, and *Nkx2-5*$^{+/-}$/*CnA*-Tg models, respectively. (c) Quantitative RT-PCR results showing ventricular mRNA expression patterns of *Egr1*, *Nab1*, *Nab2*, and *Nkx2-5* in *calcineurin* overexpression-induced cardiac hypertrophy and *Nkx2-5* haploinsufficiency. Error bars indicate s.e.m. (*, $P < 0.05$, $n = 4$).

Baseline vs. TAC and CnA-Tg

GO Accession	Population fraction	Study fraction	Description	P-value
0006817	53/7013	10/100	phosphate transport	1.6E-07
0007155	266/7013	17/100	cell adhesion	4.9E-06
0014037	3/7013	2/100	Schwann cell differentiation	1.2E-02
0030162	4/7013	2/100	regulation of proteolysis	1.8E-02
0001558	45/7013	4/100	regulation of cell growth	4.4E-02

Nkx2-5 WT vs. Nkx2-5+/-

GO Accession	Population fraction	Study fraction	Description	P-value
0008152	254/7013	17/122	metabolism	9.0E-05
0006099	20/7013	5/122	tricarboxylic acid cycle	5.3E-04
0006631	44/7013	5/122	fatty acid metabolism	8.8E-03
0006633	24/7013	4/122	fatty acid biosynthesis	1.0E-02
0019439	3/7013	2/122	aromatic compound catabolism	1.0E-02
0007031	10/7013	3/122	peroxisome organization and biogenesis	1.1E-02
0001662	4/7013	2/122	behavioral fear response	1.4E-02
0008542	5/7013	2/122	visual learning	2.1E-02
0001676	6/7013	2/122	long-chain fatty acid metabolism	2.7E-02
0006725	7/7013	2/122	aromatic compound metabolism	3.1E-02
0009083	7/7013	2/122	branched chain family amino acid catabolism	3.4E-02
0009636	9/7013	2/122	response to toxin	3.8E-02
0042593	9/7013	2/122	glucose homeostasis	4.1E-02
0006629	103/7013	6/122	lipid metabolism	4.2E-02
0006637	9/7013	2/122	acyl-CoA metabolism	4.4E-02
0006749	10/7013	2/122	glutathione metabolism	4.4E-02

Table 3.2: Gene Ontology (GO) terms associated with differential gene expression in cardiac hypertrophy and *Nkx2-5* haploinsufficiency.

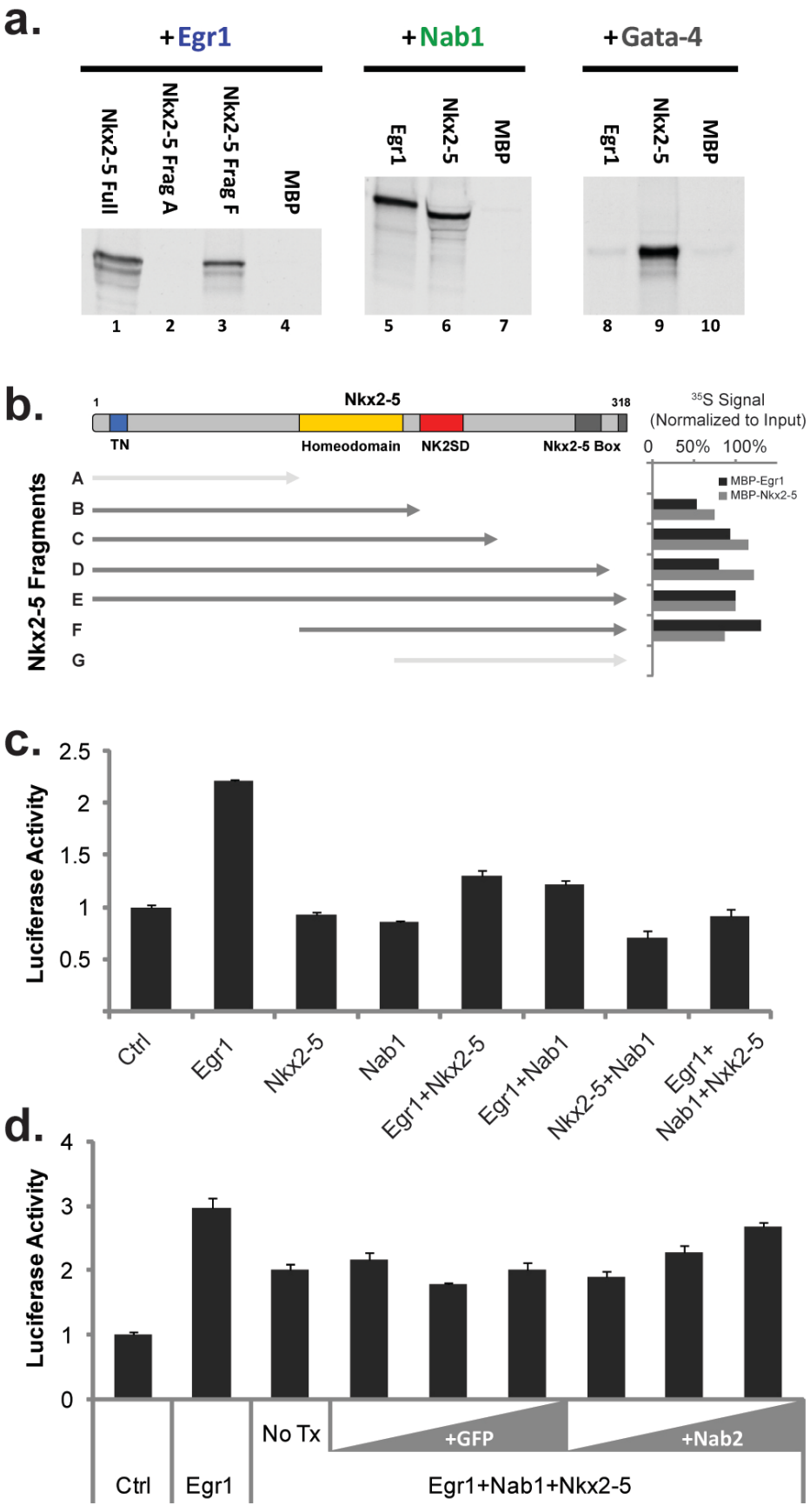


Figure 3.3: Nkx2-5 interacts with Egr1 and Nab1 to modulate Egr1-dependent

transcriptional activation in the proximal promoter of candidate target gene *Nab2*. (a) *In vitro* binding assay showing ^{35}S-labeled Egr1 physically interacting with immobilized full length Nkx2-5 (lane 1) and a C-terminal fragment of Nkx2-5 (fragment F, aa137-318, lane 3) but not a N-terminal fragment (fragment A, aa1-136). ^{35}S-labeled Nab1 also physically interacted with MBP-Nkx2-5 (lane 6). Previously described Egr1- Nab1 (lane 5) and Nkx2-5-Gata-4 (lane 9) interactions were included as controls. (b) ^{35}S-labeled truncation fragments of Nkx2-5 were tested for interaction with MBP-Egr1 or MBP-Nkx2-5 to determine the Nkx2-5 protein domain involved in these protein-protein interactions. The fragments are shown as grey arrows below a schematic drawing of Nkx2-5 structure. The darker shade indicates Nkx2-5 fragments in which MBP-Egr1 binding activity is preserved. The ^{35}S signals retained from positive interactions to MBP-Egr1 and MBP-Nkx2-5 were normalized to the input signals and shown on the right. (c) A transactivation assay demonstrating the effect of Nab1 and Nkx2-5 on Egr1-dependent activation of the *Nab2* proximal promoter (-2000-+500 relative to annotated transcriptional start site). The addition of Egr1 alone activated the *Nab2* promoter. Nkx2-5 and Nab1 both had a suppressive effect on Egr1-dependent transcriptional activation of the *Nab2* promoter. (d) A transactivation assay demonstrating the effect of Nab2 on the Nab1/Nkx2-5-dependent suppression of Egr1-dependent activation of the *Nab2* promoter. The addition of an increasing concentration of Nab2 reduced the suppressive effect of Nab1 and Nkx2-5 on Egr1-dependent activation of the *Nab2* promoter, but the addition of an increasing concentration of GFP did not. For all panels, error bars indicate s.e.m. (*, $P < 0.05$ compared with luciferase activity when only Egr1 was present, unless otherwise indicated, $n = 4$).

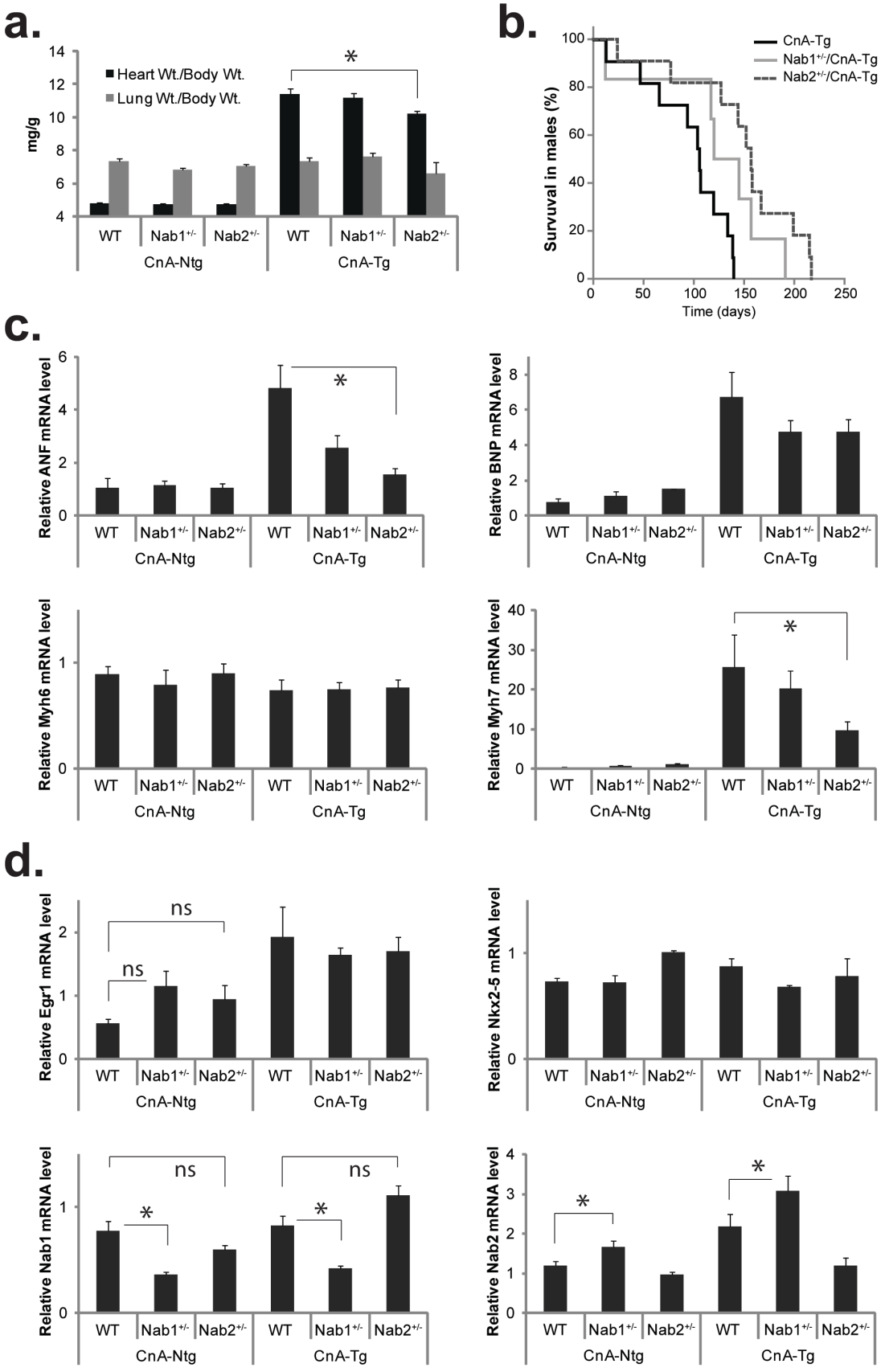


Figure 3.4: Reduced *Nab2* gene dosage modifies the phenotype in the *CnA*-Tg model

of cardiac hypertrophy. (a) *Calcineurin* overexpression-induced cardiac hypertrophy was attenuated in *Nab2*$^{+/-}$ mice. (P = 0.05, *n* = 6-11). *Nab1* or *Nab2* haploinsufficiency did not affect the cardiac mass under baseline conditions. Lung weight was unaffected by changes in *Nab1* or *Nab2* level under baseline or hypertrophic conditions. (b) A Kaplan-Meier plot showing improved survival in *Nab2*$^{+/-}$ animals compared with wild-type (WT) and *Nab1*$^{+/-}$ littermates under hypertrophic conditions (P = 0.02 and 0.03, respectively, *n* = 6-11). (c) Quantitative RT-PCR results showing the ventricular mRNA expression patterns of cardiac hypertrophy markers *ANF*, *BNP*, *Myh6*, and *Myh7* under baseline or hypertrophic conditions in WT, *Nab1*$^{+/-}$, and *Nab2*$^{+/-}$ animals. Error bars indicate s.e.m. (*, P < 0.05, *n* = 4). (d) Quantitative RT-PCR results showing the ventricular mRNA expression patterns of transcriptional regulators *Egr1*, *Nab1*, *Nab2*, and *Nkx2-5* under baseline or hypertrophic conditions in WT, *Nab1*$^{+/-}$, and *Nab2*$^{+/-}$ animals. Error bars indicate s.e.m. (*, P < 0.05; ns, not significant; *n* = 4).

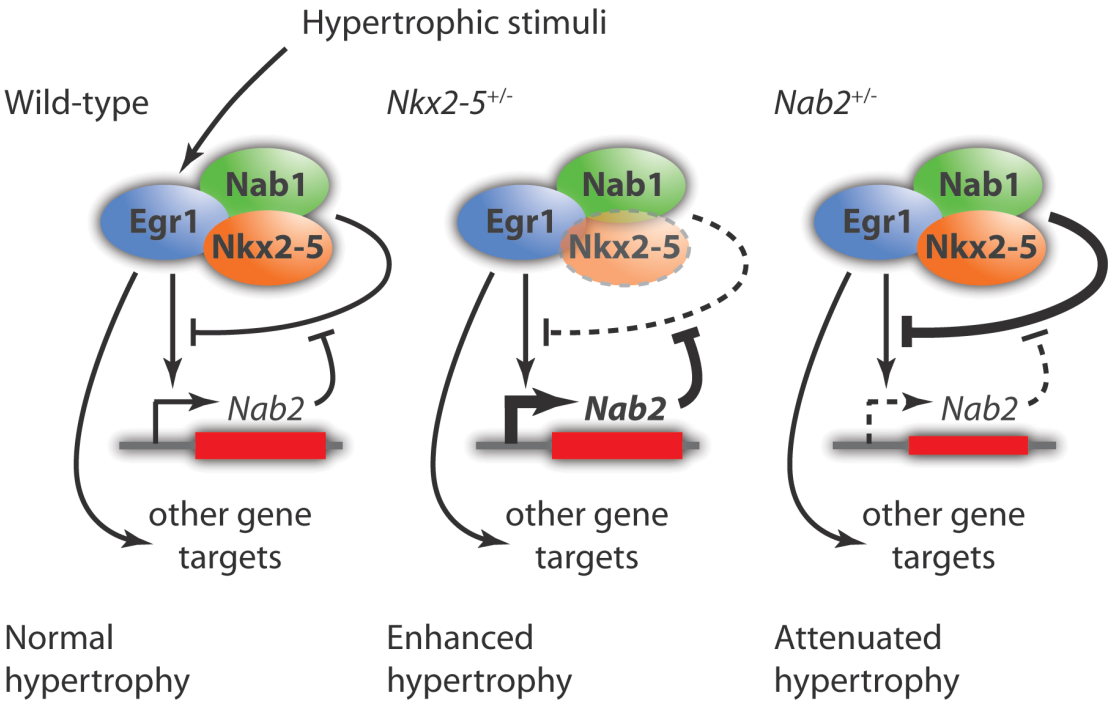


Figure 3.5: A model outlining the proposed mechanism underlying the phenotype of *Nkx2-5*$^{+/-}$ and *Nab2*$^{+/-}$ animals in pathological cardiac hypertrophy.

Supplementary Table S3.1: (Attached) Differential gene expression in mouse models of cardiac hypertrophy and *Nkx2-5* haploinsufficiency. Separate tables corresponding to the Venn diagrams shown in Figure 3.2a (left) are provided.

Supplementary Table S3.2: (Attached) Results from the TFBS screens. Separate tables corresponding to the Venn diagrams shown in Figure 3.2a (right) are provided. The 'TF_ACCESSION', 'TF_ID', and 'TF_NAME' fields contain the official accession numbers, IDs, and names for each TFBS in the TRANSFAC dataset. 'Maximum Score' field contains the best possible scores for each TFBS under the scoring model used. The 'Best Z Score', 'PV', and 'Rank' fields contain the ranksum z-scores, the corresponding p-values, and the ranks (lower is more significant) for each TFBS within the models compared, using the optimal scoring threshold recorded

in the ‘Scoring Threshold’ field to determine the presence of the particular TFBS in the proximal promoters. Repeated TFBS entries with higher ranks were removed to render the TFBS records unique.

qRT-PCR:

Probes	Forward/Reverse
Nkx2-5	AGTGCTCTCCTGCTTTCC/CTTTGTCCAGCTCCACTG
Actb	GCGTGACATCAAAGAGAAGC/AGGATTCCATACCCAAGAAGG
Nppa	GCAGAGACAGCAAACATCAG/TCTAGCAGGTTCTTGAAATCCA
Nppb	CAGAGACAGCTCTTGAAGGA/AACAACTTCAGTGCGTTACAG
Egr1	TCGGCTCCTTTCCTCACTCA/CTCATAGGGTTGTTCGCTCGG
Nab1	CTGATGGACAAGGAGAAAGAC/TTTAGAATCCCAAGGCTCTC
Myh6	GCCCAGTACCTCCGAAAGTC/GCCTTAACATACTCCTCCTTGTC
Myh7	ACTGTCAACACTAAGAGGGTCA/TTGGATGATTTGATCTTCCAGGG
Nab2	TGGCAGAGGGGATAACACAC/CGTTGCAGGACCCGATACAG

Genotyping:

Nab1 Genotype WT For	GGTCCATGTATTTTGCATATTTCTTATTT
Nab1 Genotype KO For	GCATCACAAATTTCACAAATAAAGCATTTT
Nab1 Genotype Com Rev	CACAGACTTGATGTGATTTCAGTCATTCTT
Nab2 Genotype 9598	CTTCTCGAGCACTGCCCGCCTTTTCTC
Nab2 Genotype 9599	GGAATTCGCCATGGCACTGCCTC
Nab2 Genotype 9600	ACGGGCGTCCGCGTCGTTGTTTG
Nab2 Genotype 9601	CTTCGCCCAATAGCAGCCAGTCC
Nkx2-5 PB Forward?	GACAAAGCCGAGACGGATGG
Nkx2-5 PB Reverse?	CTGTCGCTTGCACTTGTAGC
CnA genotyping forward	
CnA genotyping reverse	

Nab2 Cloning:

Nab2 Genomic PCR -2097	AGCCCCGCGCTAACAGGTTTAC
Nab2 Genomic PCR +515	TCTTCTCCCCCTGCTCCAACATC
Nab2 Promoter Cloning For	AAAACTCGAGTTCTAACATTCAAACTAACGGCTCGGCGCGG
Nab2 Promoter Cloning Rev	AAAAAAGCTTGAGTGGCGACAACTTCTCTCCAAGTTCGAGG

Nkx2-5 Fragments

Csx pcDNA Cloning 1F	AAA-GAATTC-GACACC-ATG-ttccccagccctgcgctcacacccac
Csx pcDNA Cloning 137F	AAA-GAATTC-GACACC-ATG-cgacggaagccacgcgtgctcttctcgcagg
Csx pcDNA Cloning 208F	AAA-GAATTC-GACACC-ATG-caggaccagactctggagcttctggggccg
Csx pcDNA Cloning 251F	AAA-GAATTC-GACACC-ATG-taccccagctacggcggcgcggcctgcag

Csx pcDNA Cloning 136R	AAATCTAGACTACCGCCGTGCGCGTGGTCTCTCGG
Csx pcDNA Cloning 207R	AAATCTAGACTACGGCGGCCCCAGAAGCTCCAGAGTC
Csx pcDNA Cloning 255R	AAATCTAGACTAGGGGTAGGCGTTGTAGCCATAGGCA
Csx pcDNA Cloning 313R	AAATCTAGACTAGTGCAGCGTGGAGACGCCCGAATTG
Csx pcDNA Cloning 318R	AAATCTAGACTACCAGGCTCGGATGCCGTGCAGCGTG

Supplementary Table S3.3: A list of primers used in the genotyping, cloning, and quantitative RT-PCR experiments described in the present study.

Chapter 4. Summary and Conclusion

Transcription is a fundamental mechanism underlying the causality of an organism's inherent genotype and its observed phenotype. Thus, study of transcriptional regulation forms the basis of modern molecular genetics. In development, tissue destined to make up each organ must express an appropriate set of genes which ultimately actualizes its form and function. This process is highly regulated through an intricate interplay between transcription factors and their target sequences in the genomic DNA. Transcriptional regulation is dynamic. Network of signaling proteins forms pathways directing the transcription factors to differentially regulate their target genes in response to normal or pathologic extrinsic stimuli. Understanding transcriptional regulation of a biological process is a critical step to gain the ability to modify the process and influence the outcome. As such, biomedical research efforts of the past few decades placed a major emphasis on prediction and linkage of novel transcriptional regulatory pathways to phenotype changes in development and disease. The advancement of high-throughput experimental methodology and computational biology offers unprecedented speed and ease to capture undiscovered regulatory pathways for detailed study in the context of specific biological processes. The work presented in this dissertation provides a computational framework to analyze datasets generated by high-throughput experimental methods.

The analytical framework described here consists of two complementary approaches based on rank-sum statistics to identify such pathways. The first approach enables meta-analysis of large microarray datasets encompassing thousands of samples to predict tissue-specific gene expression patterns. Each gene's expression pattern is summarized as a series of expression ranks in various tissues, thus making

tests of correlation to a phenotype of interest possible. This part of the analytical framework was applied to study various aspects of adipose tissue biology and obesity. The second approach correlates transcription factors to a transition in phenotype by detecting improbable relative abundance of specific transcription factor binding sites in the proximal promoters of genes differentially regulated in response to such transition. I applied this approach to study the modification of cardiac hypertrophy phenotype in mouse models by cardiac-specific transcription factor Nkx2-5. The computational methods set forth herein establish a flexible and unbiased mean of orthogonal data integration and rapid hypothesis generation.

In addition to the development of computational methods, I have identified and characterized novel transcriptional regulators each of which represents a potential point of therapeutic intervention in major human diseases. In the preliminary study leading to the part of my thesis research pertaining to the induction of brown fat phenotype in WAT, my analytical framework furnished a focused and testable hypothesis that nuclear receptor Esrrg and a host of putative Esrrg targets share similar tissue-specific expression pattern as BAT master regulator Prdm16. This was achieved using only open-access public data at nearly no cost to our laboratory. In the validation experiments, I showed that Esrrg indeed functions to up-regulate BAT markers such as Ucp1 in adipocyte cell culture system. Specific inhibition of Esrrg with 4-OHT, an activated tamoxifen derivative, prevented the up-regulation of Ucp1 in the WAT of mice subjected to long-term cold temperature exposure. I further identified Srebp1c as a co-regulator of Esrrg. Transfection of Srebp1c in Esrrg-rich myoblast cell line C2C12 results in strong up-regulation of BAT marker Ucp1. The addition of 4-OHT or co-transfection with vectors expressing Esrrg-targeting shRNA

in the same cell culture model attenuated Srebp1c-dependent Ucp1 up-regulation. Targeted delivery of Esrrg agonist may represent an attractive therapeutic strategy given the high interest in novel pharmaceutical treatments of obesity.

My thesis research also made progress towards a more complete understanding of the pathogenic process in cardiac hypertrophy. By combining orthogonal datasets including expression profiles from multiple mouse models of cardiac hypertrophy, genomic sequences, and experimentally determined transcription factor profiles, my computational method predicted that the Egr1 pathway may be sensitive to Nkx2-5 dosage which would account for the exacerbated cardiac hypertrophic response to detrimental stimuli observed in the Nkx2-5 haploinsufficient mice. This hypothesis was validated in vitro where a novel protein-protein interaction between Nkx2-5 and Egr1 was demonstrated. The functional consequence of this interaction appears to be the suppression of Egr1-dependent transcriptional activation. A target of this regulation is *Nab2*. Nab2 is a paralog of Egr1-specific repressor Nab1. Egr1-driven expression of Nab2 can interfere with Nab1-dependent suppression of Egr1, thus forming a feed forward loop. Indeed, male mice haploinsufficient for Nab2 have smaller hearts and improved survival in a yearlong validation study in vivo. While in no way challenging the value and validity of traditional gain- and loss-of function animal studies, my results suggest that small deviations in gene dosage can in fact lead to dramatic phenotypic differences even in complex multigenic diseases. Another unexpected, yet intriguing, discovery was the sexually dimorphic response to Nab gene dosage in the context of cardiac hypertrophy. While the male *Nab* mice behaved consistently with the predictions *in silico* and the results *in vitro*, the females did not. My results suggest that *Nab2* is expressed at a higher level in the males at baseline

and contributes more towards the development of cardiac hypertrophy in the males compared to females. This may represent a first instance of sexually dimorphic mechanism at the level of transcriptional regulation in cardiac hypertrophy that does not involve sex hormones.

Chapter 5. Future Directions

Several studies making novel use of tissue-specific gene lists are ongoing to accelerate discoveries and are generating a significant interest in this resource. In obesity, the white adipose tissue up-regulates specific adipokines, cytokines, and chemokines [12]. Using *ob* mice as a model of obesity [106], I was able to correlate the differential regulation of chemokines in the obese WAT to specific immune cell targets by including the chemokine-receptor pairing information in the analysis. The results from this study may enhance our understanding of the cell-cell communication important in the pathogenesis of obesity. The interception of such communication may lead to a novel strategy to curtail the development of obesity.

Another on-going project is the linkage analysis of atrial septal defect (ASD) in *Nkx2-5*$^{+/-}$ mice. Using microsatellite markers in a GWAS study, Julia Honold et al. were able to identify an allele that co-segregates with the development of ASD in *Nkx2-5*$^{+/-}$ neonates. [31] A typical allele resolved by microsatellite markers contains hundreds of known genes. Fine mapping, which is both cost and labor intensive, is therefore necessary to narrow down the causal gene within the genomic region. In a novel approach to this problem, we used the cardiac-specific gene list to prioritize candidate genes assuming that genes with highly restricted expression would likely be functionally important in the tissues they express in. We are currently focusing on *sarcospan* as a candidate modifier gene of *Nkx2-5* in the context of ASD development. *Sarcospan* expression is highly specific in the heart, and plays a role in the development of muscular dystrophy [107,108]. The preliminary results are near significant and encouraging. If successful, this approach can be adopted in any GWAS study where hypothesis can be reasonably made on the tissue-specificity of the causal gene.

Tissue specific gene lists can be useful in a scenario where one tissue takes on the function of another in a well-defined manner as exemplified in WAT to BAT transdifferentiation. Bielinska et al. have described development of adrenal tumors that produce sex hormones in gonadectomized mice [109]. Transcription factor *Gata4* is involved in the tumor development but the precise mechanism of tumorigenesis is not known. RNA-seq experiment to survey gene expression in the adrenal tumors is planned to gain further understanding of the tumorigenic process. The gonad-specific gene list can be applied here to facilitate the interpretation of the differential gene regulation associated with tumor transformation and sex hormone production.

At the conclusion of the study described in Chapter 2, several topics were left unexplored and warrant further investigation. Our understanding of how promoter sequences encode instruction for tissue-specific expression is incomplete. At one point in my thesis research, I had attempted to correlate global tissue-specific gene expression to combinations of TFBSs found over-represented in the promoters of a training set of known tissue-specific genes. This approach had a limited yield since the TFBS input was not reliable. Revisiting this problem using data from the latest methods such as ChIP-seq may boost probability of success. A few issues also remain unresolved regarding the validation experiments. The study did not make clear whether the forced expression of BAT markers had any functional consequence *in vitro* and *in vivo*. Further metabolic and gene expression studies could help determining what the white adipose tissue expressing BAT markers actually represent functionally. Additionally, how the Esrrg pathway relates to the Prdm16 pathway is not known. Prdm16 expression was marginally up-regulated in the WAT following

exposure to cold and was not diminished in response to addition of 4-OHT. Prdm16 may be necessary but insufficient in cold-induced BAT gene activation in WAT. Direct testing of Prdm16 in this context is needed to better define its functional relation to Esrrg.

My work on Nkx2-5 haploinsufficiency described in Chapter 3 similarly opens opportunities for further study. While my results point to a model where Nkx2-5 haploinsufficiency modifies target gene expressions through changes in the dynamics of physical interaction between Nkx2-5 and transcriptional co-regulators, they do not rule out a mechanism where Nkx2-5-DNA interactions are altered directly to cause differential gene regulation. To this end, I had unsuccessfully analyzed the proximal promoter sequences in differentially regulated genes in the *Nkx2-5*$^{+/-}$ hearts compared to wild-type hearts under non-hypertrophic conditions. Nkx2-5 ChIP data and more accurate and comprehensive surveying of global mRNA expression have since become available which may facilitate the discovery of promoter/enhancer sequence elements that specify *Nkx2-5* dosage sensitivity.

The conclusions from the series of animal studies carried out for this work may be bolstered by additional experiments. Due to the availability of mice and time and space constraints I was not able to directly characterize mice with reduced Egr1 expression levels in the context of calcineurin overexpression induced cardiac hypertrophy. Similarly, how the *Nab* haploinsufficiency interacts with Nkx2-5 haploinsufficiency in the context of cardiac hypertrophy has not been investigated in full. I hypothesize that, at least in the males, Nab2 haploinsufficiency may attenuate the enhanced hypertrophic response in the *Nkx2-5*$^{+/-}$/*CnA*-Tg mice. Mice needed to

test this hypothesis currently are being bred in the Jay lab. In addition, the *Nab* haploinsufficient mice may be further characterized in a different model of cardiac hypertrophy such as the TAC model. Additional phenotyping experiments such as echocardiography and electrophysiology studies may be useful in establishing the functional consequences of Nab haploinsufficiency in the context of cardiac hypertrophy.

References

1. Flegal, K.M., Carroll, M.D., Ogden, C.L. & Curtin, L.R. Prevalence and trends in obesity among US adults, 1999-2008. *JAMA* **303**, 235-241 (2010).
2. Whitlock, G.*, et al.* Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies. *Lancet* **373**, 1083-1096 (2009).
3. Finkelstein, E.A., Trogdon, J.G., Cohen, J.W. & Dietz, W. Annual medical spending attributable to obesity: payer-and service-specific estimates. *Health Aff (Millwood)* **28**, w822-831 (2009).
4. Flegal, K.M., Carroll, M.D., Ogden, C.L. & Curtin, L.R. Prevalence and trends in obesity among US adults, 1999-2008. *Jama* **303**, 235-241 (2010).
5. Stunkard, A.J., Harris, J.R., Pedersen, N.L. & McClearn, G.E. The body-mass index of twins who have been reared apart. *N Engl J Med* **322**, 1483-1487 (1990).
6. Austin, M.A.*, et al.* Genetic influences on changes in body mass index: a longitudinal analysis of women twins. *Obes Res* **5**, 326-331 (1997).
7. Christakis, N.A. & Fowler, J.H. The spread of obesity in a large social network over 32 years. *N Engl J Med* **357**, 370-379 (2007).
8. Turnbaugh, P.J.*, et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027-1031 (2006).
9. Ley, R.E., Turnbaugh, P.J., Klein, S. & Gordon, J.I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022-1023 (2006).
10. Sackmann-Sala, L., Berryman, D.E., Munn, R.D., Lubbers, E.R. & Kopchick, J.J. Heterogeneity Among White Adipose Tissue Depots in Male C57BL/6J Mice. *Obesity (Silver Spring)* (2011).
11. Gil, A., Olza, J., Gil-Campos, M., Gomez-Llorente, C. & Aguilera, C.M. Is adipose tissue metabolically different at different sites? *Int J Pediatr Obes* **6 Suppl 1**, 13-20 (2011).
12. Hauner, H. Secretory factors from human adipose tissue and their functional role. *Proc Nutr Soc* **64**, 163-169 (2005).
13. Halaas, J.L.*, et al.* Weight-reducing effects of the plasma protein encoded by the obese gene. *Science* **269**, 543-546 (1995).
14. Ukkola, O. & Santaniemi, M. Adiponectin: a link between excess adiposity and associated comorbidities? *J Mol Med (Berl)* **80**, 696-702 (2002).
15. Weisberg, S.P.*, et al.* CCR2 modulates inflammatory and metabolic effects of high-fat feeding. *J Clin Invest* **116**, 115-124 (2006).
16. Roger, V.L.*, et al.* Heart disease and stroke statistics--2011 update: a report

from the American Heart Association. *Circulation* **123**, e18-e209 (2011).
17. Levy, D.*, et al.* Long-term trends in the incidence of and survival with heart failure. *N Engl J Med* **347**, 1397-1402 (2002).
18. Ho, K.K., Pinsky, J.L., Kannel, W.B. & Levy, D. The epidemiology of heart failure: the Framingham Study. *J Am Coll Cardiol* **22**, 6A-13A (1993).
19. Bernardo, B.C., Weeks, K.L., Pretorius, L. & McMullen, J.R. Molecular distinction between physiological and pathological cardiac hypertrophy: experimental findings and therapeutic strategies. *Pharmacol Ther* **128**, 191-227 (2010).
20. Colella, M.*, et al.* Ca2+ oscillation frequency decoding in cardiac cell hypertrophy: role of calcineurin/NFAT as Ca2+ signal integrators. *Proc Natl Acad Sci U S A* **105**, 2859-2864 (2008).
21. Molkentin, J.D.*, et al.* A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. *Cell* **93**, 215-228 (1998).
22. Lehman, J.J. & Kelly, D.P. Gene regulatory mechanisms governing energy metabolism during cardiac hypertrophic growth. *Heart Fail Rev* **7**, 175-185 (2002).
23. Oka, T., Xu, J. & Molkentin, J.D. Re-employment of developmental transcription factors in adult heart disease. *Semin Cell Dev Biol* **18**, 117-131 (2007).
24. Deshaies, Y., LeBlanc, J. & Willemot, J. Studies on protein metabolism during isoproterenol-induced cardiac hypertrophy. *Recent Adv Stud Cardiac Struct Metab* **8**, 387-395 (1975).
25. Olivetti, G.*, et al.* Apoptosis in the failing human heart. *N Engl J Med* **336**, 1131-1141 (1997).
26. Teiger, E.*, et al.* Apoptosis in pressure overload-induced heart hypertrophy in the rat. *J Clin Invest* **97**, 2891-2897 (1996).
27. Rockman, H.A.*, et al.* Segregation of atrial-specific and inducible expression of an atrial natriuretic factor transgene in an in vivo murine model of cardiac hypertrophy. *Proc Natl Acad Sci U S A* **88**, 8277-8281 (1991).
28. Regen, D.M. Calculation of left ventricular wall stress. *Circ Res* **67**, 245-252 (1990).
29. Ikeda, Y.*, et al.* Novel point mutation in the cardiac transcription factor CSX/NKX2.5 associated with congenital heart disease. *Circ J* **66**, 561-563 (2002).
30. Schott, J.J.*, et al.* Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* **281**, 108-111 (1998).
31. Winston, J.B.*, et al.* Heterogeneity of genetic modifiers ensures normal

cardiac development. *Circulation* **121**, 1313-1321 (2010).

32. Komuro, I. & Izumo, S. Csx: a murine homeobox-containing gene specifically expressed in the developing heart. *Proc Natl Acad Sci U S A* **90**, 8145-8149 (1993).
33. Saadane, N., Alpert, L. & Chalifour, L.E. Expression of immediate early genes, GATA-4, and Nkx-2.5 in adrenergic-induced cardiac hypertrophy and during regression in adult mice. *Br J Pharmacol* **127**, 1165-1176 (1999).
34. Toko, H.*, et al.* Csx/Nkx2-5 is required for homeostasis and survival of cardiac myocytes in the adult heart. *J Biol Chem* **277**, 24735-24743 (2002).
35. Wakimoto, H., Kasahara, H., Maguire, C.T., Izumo, S. & Berul, C.I. Developmentally modulated cardiac conduction failure in transgenic mice with fetal or postnatal overexpression of DNA nonbinding mutant Nkx2.5. *J Cardiovasc Electrophysiol* **13**, 682-688 (2002).
36. Takeda, M.*, et al.* Slow progressive conduction and contraction defects in loss of Nkx2-5 mice after cardiomyocyte terminal differentiation. *Lab Invest* **89**, 983-993 (2009).
37. Kasahara, H.*, et al.* Nkx2.5 homeoprotein regulates expression of gap junction protein connexin 43 and sarcomere organization in postnatal cardiomyocytes. *J Mol Cell Cardiol* **35**, 243-256 (2003).
38. Takimoto, E.*, et al.* Up-regulation of natriuretic peptides in the ventricle of Csx/Nkx2-5 transgenic mice. *Biochem Biophys Res Commun* **270**, 1074-1079 (2000).
39. Edgar, R., Domrachev, M. & Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-210 (2002).
40. Barrett, T.*, et al.* NCBI GEO: archive for functional genomics data sets--10 years on. *Nucleic Acids Res* **39**, D1005-1010 (2011).
41. Leek, J.T.*, et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* **11**, 733-739 (2010).
42. Jorgensen, E.M. & Mango, S.E. The art and design of genetic screens: caenorhabditis elegans. *Nat Rev Genet* **3**, 356-369 (2002).
43. Ioannidis, J.P.*, et al.* Repeatability of published microarray gene expression analyses. *Nat Genet* **41**, 149-155 (2009).
44. Shi, L.*, et al.* The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat Biotechnol* **28**, 827-838 (2010).
45. Kawamoto, S.*, et al.* BodyMap: a collection of 3' ESTs for analysis of human gene expression information. *Genome Res* **10**, 1817-1827 (2000).

46. Kouadjo, K.E., Nishida, Y., Cadrin-Girard, J.F., Yoshioka, M. & St-Amand, J. Housekeeping and tissue-specific genes in mouse tissues. *BMC Genomics* **8**, 127 (2007).
47. Kumbrink, J., Gerlinger, M. & Johnson, J.P. Egr-1 induces the expression of its corepressor nab2 by activation of the nab2 promoter thereby establishing a negative feedback loop. *J Biol Chem* **280**, 42785-42793 (2005).
48. Hsiao, L.L.*, et al.* A compendium of gene expression in normal human tissues. *Physiol Genomics* **7**, 97-104 (2001).
49. Su, A.I.*, et al.* Large-scale analysis of the human and mouse transcriptomes. *Proc Natl Acad Sci U S A* **99**, 4465-4470 (2002).
50. Su, A.I.*, et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* **101**, 6062-6067 (2004).
51. Zhang, W.*, et al.* The functional landscape of mouse gene expression. *J Biol* **3**, 21 (2004).
52. Cannon, B., Hedin, A. & Nedergaard, J. Exclusive occurrence of thermogenin antigen in brown adipose tissue. *FEBS Lett* **150**, 129-132 (1982).
53. Cousin, B.*, et al.* Occurrence of brown adipocytes in rat white adipose tissue: molecular and morphological characterization. *J Cell Sci* **103 (Pt 4)**, 931-942 (1992).
54. Barbatelli, G.*, et al.* The emergence of cold-induced brown adipocytes in mouse white fat depots is determined predominantly by white to brown adipocyte transdifferentiation. *Am J Physiol Endocrinol Metab* **298**, E1244-1253 (2010).
55. Kajimura, S.*, et al.* Initiation of myoblast to brown fat switch by a PRDM16-C/EBP-beta transcriptional complex. *Nature* **460**, 1154-1158 (2009).
56. Kajimura, S.*, et al.* Regulation of the brown and white fat gene programs through a PRDM16/CtBP transcriptional complex. *Genes Dev* **22**, 1397-1409 (2008).
57. Seale, P.*, et al.* PRDM16 controls a brown fat/skeletal muscle switch. *Nature* **454**, 961-967 (2008).
58. Seale, P.*, et al.* Transcriptional control of brown fat determination by PRDM16. *Cell Metab* **6**, 38-54 (2007).
59. Seale, P.*, et al.* Prdm16 determines the thermogenic program of subcutaneous white adipose tissue in mice. *J Clin Invest* **121**, 96-105 (2010).
60. Kajimura, S., Seale, P. & Spiegelman, B.M. Transcriptional control of brown fat development. *Cell Metab* **11**, 257-262 (2010).
61. Minty, A.J., Alonso, S., Caravatti, M. & Buckingham, M.E. A fetal skeletal

muscle actin mRNA in the mouse and its identity with cardiac actin mRNA. *Cell* **30**, 185-192 (1982).

62. Leonard, A.E.*, et al.* Cloning of a human cDNA encoding a novel enzyme involved in the elongation of long-chain polyunsaturated fatty acids. *Biochem J* **350**, 765-770 (2000).
63. Akamine, R.*, et al.* Usefulness of the 5' region of the cDNA encoding acidic ribosomal phosphoprotein P0 conserved among rats, mice, and humans as a standard probe for gene expression analysis in different tissues and animal species. *J Biochem Biophys Methods* **70**, 481-486 (2007).
64. Vierbuchen, T.*, et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035-1041 (2010).
65. Alaynick, W.A.*, et al.* ERRgamma directs and maintains the transition to oxidative metabolism in the postnatal heart. *Cell Metab* **6**, 13-24 (2007).
66. Wolins, N.E.*, et al.* OP9 mouse stromal cells rapidly differentiate into adipocytes: characterization of a useful new model of adipogenesis. *J Lipid Res* **47**, 450-460 (2006).
67. Coward, P., Lee, D., Hull, M.V. & Lehmann, J.M. 4-Hydroxytamoxifen binds to and deactivates the estrogen-related receptor gamma. *Proc Natl Acad Sci U S A* **98**, 8880-8884 (2001).
68. Greschik, H., Flaig, R., Renaud, J.P. & Moras, D. Structural basis for the deactivation of the estrogen-related receptor gamma by diethylstilbestrol or 4-hydroxytamoxifen and determinants of selectivity. *J Biol Chem* **279**, 33639-33646 (2004).
69. Shimomura, I., Shimano, H., Horton, J.D., Goldstein, J.L. & Brown, M.S. Differential expression of exons 1a and 1c in mRNAs for sterol regulatory element binding protein-1 in human and mouse organs and cultured cells. *J Clin Invest* **99**, 838-845 (1997).
70. Mecham, B.H., Nelson, P.S. & Storey, J.D. Supervised normalization of microarrays. *Bioinformatics* **26**, 1308-1315 (2010).
71. Puigserver, P.*, et al.* A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell* **92**, 829-839 (1998).
72. Wu, Z.*, et al.* Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. *Cell* **98**, 115-124 (1999).
73. Villanueva, C.J.*, et al.* TLE3 Is a Dual-Function Transcriptional Coregulator of Adipogenesis. *Cell Metab* **13**, 413-427 (2011).
74. Castillo-Davis, C.I. & Hartl, D.L. GeneMerge--post-genomic analysis, data mining, and hypothesis testing. *Bioinformatics* **19**, 891-892 (2003).

75. Marshall, O.J. PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics* **20**, 2471-2472 (2004).
76. Spandidos, A., Wang, X., Wang, H. & Seed, B. PrimerBank: a resource of human and mouse PCR primer pairs for gene expression detection and quantification. *Nucleic Acids Res* **38**, D792-799 (2010).
77. Berenji, K., Drazner, M.H., Rothermel, B.A. & Hill, J.A. Does load-induced ventricular hypertrophy progress to systolic heart failure? *Am J Physiol Heart Circ Physiol* **289**, H8-H16 (2005).
78. Hoshijima, M. & Chien, K.R. Mixed signals in heart failure: cancer rules. *J Clin Invest* **109**, 849-855 (2002).
79. Tanaka, M., Chen, Z., Bartunkova, S., Yamasaki, N. & Izumo, S. The cardiac homeobox gene Csx/Nkx2.5 lies genetically upstream of multiple genes essential for heart development. *Development* **126**, 1269-1280 (1999).
80. Lyons, I.*, et al.* Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes Dev* **9**, 1654-1666 (1995).
81. Jay, P.Y.*, et al.* Haploinsufficiency of the cardiac transcription factor Nkx2-5 variably affects the expression of putative target genes. *FASEB J* **19**, 1495-1497 (2005).
82. Briggs, L.E.*, et al.* Perinatal loss of Nkx2-5 results in rapid conduction and contraction defects. *Circ Res* **103**, 580-590 (2008).
83. Jay, P.Y.*, et al.* Nkx2-5 mutation causes anatomic hypoplasia of the cardiac conduction system. *J Clin Invest* **113**, 1130-1137 (2004).
84. Tarnavski, O.*, et al.* Mouse cardiac surgery: comprehensive techniques for the generation of mouse models of human diseases and their application for genomic studies. *Physiol Genomics* **16**, 349-360 (2004).
85. Satoh, H., Delbridge, L.M., Blatter, L.A. & Bers, D.M. Surface:volume relationship in cardiac myocytes studied with confocal microscopy and membrane capacitance measurements: species-dependence and developmental effects. *Biophys J* **70**, 1494-1504 (1996).
86. Dong, D.*, et al.* Overexpression of calcineurin in mouse causes sudden cardiac death associated with decreased density of K+ channels. *Cardiovasc Res* **57**, 320-332 (2003).
87. Zhao, M., Chow, A., Powers, J., Fajardo, G. & Bernstein, D. Microarray analysis of gene expression after transverse aortic constriction in mice. *Physiol Genomics* **19**, 93-105 (2004).
88. Wingender, E., Dietze, P., Karas, H. & Knuppel, R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res* **24**,

238-241 (1996).

89. Saadane, N., Alpert, L. & Chalifour, L.E. Altered molecular response to adrenoreceptor-induced cardiac hypertrophy in Egr-1-deficient mice. *Am J Physiol Heart Circ Physiol* **278**, H796-805 (2000).

90. Brand, T., Sharma, H.S. & Schaper, W. Expression of nuclear proto-oncogenes in isoproterenol-induced cardiac hypertrophy. *J Mol Cell Cardiol* **25**, 1325-1337 (1993).

91. Iwaki, K., Sukhatme, V.P., Shubeita, H.E. & Chien, K.R. Alpha- and beta-adrenergic stimulation induces distinct patterns of immediate early gene expression in neonatal rat myocardial cells. fos/jun expression is associated with sarcomere assembly; Egr-1 induction is primarily an alpha 1-mediated response. *J Biol Chem* **265**, 13809-13817 (1990).

92. Buitrago, M.*, et al.* The transcriptional repressor Nab1 is a specific regulator of pathological cardiac hypertrophy. *Nat Med* **11**, 837-844 (2005).

93. Svaren, J.*, et al.* NAB2, a corepressor of NGFI-A (Egr-1) and Krox20, is induced by proliferative and differentiative stimuli. *Mol Cell Biol* **16**, 3545-3553 (1996).

94. Russo, M.W., Sevetson, B.R. & Milbrandt, J. Identification of NAB1, a repressor of NGFI-A- and Krox20-mediated transcription. *Proc Natl Acad Sci U S A* **92**, 6873-6877 (1995).

95. Lee, Y.*, et al.* The cardiac tissue-restricted homeobox protein Csx/Nkx2.5 physically associates with the zinc finger protein GATA4 and cooperatively activates atrial natriuretic factor gene expression. *Mol Cell Biol* **18**, 3120-3129 (1998).

96. Kasahara, H.*, et al.* Characterization of homo- and heterodimerization of cardiac Csx/Nkx2.5 homeoprotein. *J Biol Chem* **276**, 4570-4580 (2001).

97. Bisping, E.*, et al.* Gata4 is required for maintenance of postnatal cardiac function and protection from pressure overload-induced heart failure. *Proc Natl Acad Sci U S A* **103**, 14471-14476 (2006).

98. Svaren, J.*, et al.* NAB2, a corepressor of NGFI-A (Egr-1) and Krox20, is induced by proliferative and differentiative stimuli. *Mol Cell Biol* **16**, 3545-3553 (1996).

99. Le, N.*, et al.* Nab proteins are essential for peripheral nervous system myelination. *Nat Neurosci* **8**, 932-940 (2005).

100. Mootha, V.K.*, et al.* Erralpha and Gabpa/b specify PGC-1alpha-dependent oxidative phosphorylation gene expression that is altered in diabetic muscle. *Proc Natl Acad Sci U S A* **101**, 6570-6575 (2004).

101. Hertz, G.Z. & Stormo, G.D. Identifying DNA and protein patterns with

statistically significant alignments of multiple sequences. *Bioinformatics* **15**, 563-577 (1999).

102. Nishida, K., Frith, M.C. & Nakai, K. Pseudocounts for transcription factor binding sites. *Nucleic Acids Res* **37**, 939-944 (2009).
103. Staden, R. Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Res* **12**, 505-519 (1984).
104. Brunet, S.*, et al.* Heterogeneous expression of repolarizing, voltage-gated K+ currents in adult mouse ventricles. *J Physiol* **559**, 103-120 (2004).
105. Xu, A., Netticadan, T., Jones, D.L. & Narayanan, N. Serine phosphorylation of the sarcoplasmic reticulum Ca(2+)-ATPase in the intact beating rabbit heart. *Biochem Biophys Res Commun* **264**, 241-246 (1999).
106. Ingalls, A.M., Dickie, M.M. & Snell, G.D. Obese, a new mutation in the house mouse. *J Hered* **41**, 317-318 (1950).
107. Araishi, K.*, et al.* Loss of the sarcoglycan complex and sarcospan leads to muscular dystrophy in beta-sarcoglycan-deficient mice. *Hum Mol Genet* **8**, 1589-1598 (1999).
108. Coral-Vazquez, R.*, et al.* Disruption of the sarcoglycan-sarcospan complex in vascular smooth muscle: a novel mechanism for cardiomyopathy and muscular dystrophy. *Cell* **98**, 465-474 (1999).
109. Bielinska, M.*, et al.* Gonadotropin-induced adrenocortical neoplasia in NU/J nude mice. *Endocrinology* **146**, 3975-3984 (2005).